Social Network Analysis and Mining (2025) 15:52
https://doi.org/10.1007/s13278-025-01463-6

ORIGINAL ARTICLE

Sentiment-driven cryptocurrency forecasting: analyzing LSTM, GRU, Bi-LSTM, and temporal attention model (TAM)

Phumudzo Lloyd Seabe[1] · **Claude Rodrigue Bambe Moutsinga**[1] · **Edson Pindza**[2]

Received: 29 July 2024 / Revised: 15 April 2025 / Accepted: 17 April 2025

Abstract

Predicting cryptocurrency prices is challenging due to market volatility and external influences like social media sentiment. This study integrates Twitter sentiment analysis with deep learning models (LSTM, GRU, Bi-LSTM, and Temporal Attention Model) to enhance Bitcoin price forecasting. Sentiment features were extracted using VADER and RoBERTa, with findings showing that RoBERTa-based models significantly outperform VADER. Bi-LSTM (RoBERTa) achieved the lowest MAPE of 2.01%, demonstrating the effectiveness of deep contextual embeddings. SHAP analysis identified Sentiment Momentum, RoBERTa Compound Score, and VADER Negativity Score as key predictors of price movements. These results highlight the value of sentiment-driven forecasting and provide insights for traders, investors, and researchers.

Keywords Cryptocurrency forecasting · Twitter(X) sentiment · LSTM · GRU · Bi-directional LSTM · VADER-lexicon · RoBERTa · SHapley additive explanations

1 Introduction

Cryptocurrencies are a new asset class that has come to light in recent years and drastically changed the financial scene by offering a transparent and decentralised alternative for established financial institutions. The cryptocurrency market presents possibilities and challenges for stakeholders due to its high volatility and influence from a variety of factors, such as news, investor psychology, and social media sentiment (Shen et al. 2019). Cryptocurrencies, such as Bitcoin, Ethereum, and Ripple, have not only gained substantial market capitalization but have also attracted widespread attention from investors, traders, researchers, and the general public (Nakamoto 2008). For investors looking to optimise profits and reduce risks, predicting cryptocurrency prices with accuracy has become essential to efficiently navigate this dynamic market.

✉ Phumudzo Lloyd Seabe
Phumudzo@aims.ac.za

[1] Department of Mathematics and Applied Mathematics, Sefako Makgatho Health Sciences University, Pretoria 0204, Republic of South Africa

[2] Department of Decision Sciences, College of Economic and Management Sciences, University of South Africa, Pretoria 0002, Republic of South Africa

The broad complexities of the cryptocurrency market may not be captured by traditional forecasting techniques, which frequently depend on previous price data and technical indications (Lahmiri and Bekiros 2019). However, one of the most intriguing aspects is the vulnerability of cryptocurrency trading to market sentiment, especially that exhibited on social media sites like Twitter. Twitter's real-time nature and extensive use have made it a crucial medium for sharing news, opinions, and market feelings (Arias et al. 2014). Many real-time sentiment data have become accessible with the growth of social media sites like Twitter, which may offer insightful information on the behaviour of the market (Rognone et al. 2020). Participants of cryptocurrencies frequently post their thoughts, news, and feelings on Twitter, which can affect how the market perceives things and price movements (Colianni et al. 2015). Sentiment analysis as a forecasting technique has drawn interest from researchers who have hypothesised that social media sentiments may affect cryptocurrency pricing. This concept is consistent with the Efficient Market Hypothesis (EMH), according to which asset prices represent all information that is accessible (Fama 1970). In the context of cryptocurrencies, this information often includes sentiment data from social media.

Sentiment analysis involves the extraction of subjective information from textual data, classifying it as positive, negative, or neutral. Various methodologies have been developed to perform sentiment analysis, broadly classified into rule-based approaches and machine learning techniques. Rule-based sentiment analysis relies on manually crafted rules or lexicons to identify sentiment words and their associated polarities. Although these techniques are simple and understandable, they frequently are not flexible enough to deal with changing language usage and context-dependent interpretations (Taboada et al. 2011). The Valence Aware Dictionary and Sentiment Reasoner (VADER) is one of the most successful rule-based approaches to sentiment analysis. VADER can handle textual nuances including emoticons, slang, and punctuation and is especially tuned to feelings conveyed on social media (Hutto and Gilbert 2014b). With valence scores determined by intensity and polarity, VADER provides a more detailed examination of sentiment than traditional lexicon-based techniques. In this study, VADER is further improved to better represent the sentiments surrounding bitcoin conversations on Twitter by adding a language exclusive to the cryptocurrency space. In contrast, machine learning techniques leverage large datasets to train models that can automatically discover relationships and learn patterns within the data. These techniques include supervised learning methods, where models are trained on labelled datasets, and unsupervised learning approaches, which estimate sentiment without specified labels. Popular deep learning models such Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN), as well as Naive Bayes and Support Vector Machines (SVM) are used for sentiment analysis (Pang et al. 2008).

Recurrent Neural Networks (RNNs) and their variants, including LSTM, GRU, and Bi-LSTM, are highly effective for sequential data modeling, particularly in financial time series forecasting. LSTMs address the vanishing gradient problem and capture long-term dependencies, while GRUs offer a simpler architecture with fewer parameters, enhancing computational efficiency. Bi-LSTMs further improve performance by processing data bidirectionally, enabling better recognition of sentiment patterns (Hochreiter and Schmidhuber 1997). Temporal Attention Mechanisms (TAMs) provide a robust alternative by dynamically weighting past observations to identify their relevance to future price movements, reducing noise and improving interpretability in volatile markets like cryptocurrency (Vaswani et al. 2017). Transformer-based models like BERT and GPT, while dominant in NLP, are less suited for time-series forecasting due to their quadratic computational complexity. Temporal Convolutional Networks (TCNs) are another option but require large datasets for effective training. Studies indicate that LSTMs and GRUs outperform transformers in low-data environments for financial prediction tasks due to their inherent bias toward sequential dependencies (Lim and Zohren 2021). Given these considerations, LSTM-based models combined with Temporal Attention Mechanisms are adopted for efficient and accurate time-series forecasting, leveraging attention to enhance prediction performance.

Deep learning models, particularly RNNs and their variants, have shown remarkable success in capturing the temporal dependencies and contextual nuances of textual data. Long Short-Term Memory (LSTM) networks, originally described by Hochreiter and Schmidhuber (1997), preserve long-term reliance by means of their unique cell state and gating mechanism, therefore preventing the vanishing gradient problem. This feature makes sequential data modelling and time series forecasting appropriate applications for LSTMs. Gated Recurrent Units (GRUs), proposed by Cho et al. (2014b), are a simplified variant of LSTMs that merge the forget and input gates into a single update gate, reducing the model complexity while retaining performance. Bi-Directional LSTMs extend the standard LSTM by processing the input data in both forward and backward directions, thereby capturing past and future context simultaneously (Schuster and Paliwal 1997). The application of these advanced neural network architectures to sentiment analysis for cryptocurrency forecasting represents a significant advancement in the field. By leveraging Twitter sentiment data, analyzed using VADER with an additional lexicon, and RoBERTa researchers aim to improve the accuracy and reliability of cryptocurrency price predictions. Previous studies have demonstrated the potential of sentiment analysis in predicting stock market movements and foreign exchange rates (Bollen et al. 2011). However the very erratic and speculative nature of cryptocurrencies brings special challenges which require for tailored approaches and thorough analysis.

To validate our hypotheses, we utilize real cryptocurrency price data combined with Twitter sentiment analysis. The dataset is divided into training and validation sets. Deep learning models are trained on the training data and evaluated on the validation set using multiple performance metrics, including Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Directional Accuracy (DA), to ensure a comprehensive assessment of forecasting accuracy. To assess the reliability of our models and the contribution of sentiment features in cryptocurrency price forecasting, we conduct sensitivity analyses. This study contributes to the growing body of research on leveraging social media data for financial forecasting and decision-making. By comparing the performance of deep learning architectures, we aim to provide insights into the optimal integration of Twitter sentiment data for cryptocurrency prediction, offering valuable guidance for traders and investors seeking to refine their decision-making processes and optimize trading strategies.

This research evaluates the efficacy of LSTM, GRU, Bi-LSTM, and the Temporal Attention Model (TAM) in forecasting Bitcoin prices using sentiment analysis. The study involves the collection and preprocessing of Twitter data, sentiment analysis using VADER and RoBERTa, and the implementation and comparison of deep learning models. The primary objectives are to assess the predictive power of Twitter sentiment on cryptocurrency prices and identify the most effective neural network architecture for this task.

This paper contributes to the growing body of research on sentiment-driven cryptocurrency forecasting through the following novel aspects:

1. Multi-Model Forecasting Framework: We introduce a unified sentiment-enhanced forecasting framework that systematically compares two distinct sentiment analysis approaches-VADER (a lexicon-based model) and RoBERTa (a transformer-based model)-across four deep learning architectures: LSTM, GRU, Bi-Directional LSTM, and the Temporal Attention Model (TAM).
2. Feature Innovation: We design a sentiment feature engineering pipeline that extracts temporally-aware sentiment dynamics for both VADER and RoBERTa sentiment inputs. These temporally-aware features capture short-term sentiment memory, momentum shifts, rolling variability, and sentiment-volatility interactions. By enriching both lexicon-based and deep contextual sentiment representations, this approach enhances the model's ability to detect behavioral patterns and emotional dynamics in the cryptocurrency market.
3. Explainable Forecasting with SHAP: To address the "black-box" nature of deep learning models, we incorporate SHapley Additive exPlanations (SHAP) to quantify the relative importance of sentiment and financial features. This provides valuable interpretability for traders, researchers, and developers of algorithmic trading systems.
4. Large-Scale, Long-Horizon Dataset: We employ an extensive dataset of over 48 million Bitcoin-related tweets spanning more than two years (February 2021–March 2023). This temporal scope exceeds most prior studies and enables a more robust evaluation of sentiment forecasting models under varying market conditions.
5. Empirical and Practical Insights: We empirically demonstrate that transformer-based sentiment models (RoBERTa) provide superior predictive accuracy and directional trend performance compared to lexicon-based models (VADER), highlighting their practical relevance for sentiment-aware trading and real-time market applications.

The remainder of this paper is organised as follows: Sect. 2 reviews the relevant literature on cryptocurrency forecasting and sentiment analysis. Section 3 describes the data collection and pre-processing methods, as well as the implementation of LSTM, GRU, Bi-LSTM, and Temporal Attention Models. Section 4 presents the experimental results, comparative analysis, and discusses the findings and their implications, and Sect. 5 concludes the paper with suggestions for future research.

2 Literature review

In this section, a comprehensive literature review is conducted on cryptocurrency forecasting and sentiment analysis. Firstly, we discuss the evolution and characteristics of the cryptocurrency market. Then, we explore the predictability and price discovery mechanisms within these markets. Lastly, we examine the existing literature on financial market sentiment analysis, with a focus on Twitter sentiment and its application in predicting cryptocurrency prices.

2.1 Cryptocurrencies overview

2.1.1 The market of cryptocurrency

In the late 2008, an anonymous entity known by the pseudonym Satoshi Nakamoto released a groundbreaking whitepaper introducing a decentralized cryptographic cash system. This publication laid the foundation for blockchain technology and its most renowned application, Bitcoin (Nakamoto 2008). Nakamoto's whitepaper is regarded as revolutionary because it addressed major issues in creating a secure and reliable digital currency, such as the double-spending problem, network vulnerabilities to hacking, and the inefficiencies associated with cross-border and interbank transactions (Narayanan et al. 2016; Antonopoulos 2017). Following Bitcoin's introduction, several other cryptocurrencies, commonly referred to as altcoins, emerged, including Ethereum and Litecoin. These altcoins were developed to address Bitcoin's constraints, such as its finite supply, high energy consumption, and reliance on the Proof-of-Work consensus mechanism (Vigna and Casey 2016; Buterin et al. 2014). Moreover, the introduction of smart contracts by Ethereum opened new possibilities for decentralized applications, further expanding the blockchain ecosystem (Wood et al. 2014). Initially, cryptocurrencies were often associated with illicit activities, earning them a dubious reputation as tools for criminals (Janze 2017). This negative perception was largely due to their use in black market transactions and their anonymity qualities, which attracted illegal activities like drug trafficking and money laundering, were major contributors to this unfavourable impression. However, this perception began to shift dramatically as interest in the cryptocurrency market surged in 2017 and early 2018, leading to a speculative bubble driven by fear of missing out (FOMO). During this period, the market experienced unprecedented growth, with new investors flocking to capitalize on the rapid price increases. Today, the total market capitalization of cryptocurrencies exceeds $2.77 trillion, with over 9,000 different coins in circulation (coinmarketcap.com accessed on 4 May 2024). One noteworthy instance of this speculative frenzy is described in Corbet et al. (2019), when a company's stock price increased significantly just by announcing its cryptocurrency development. This period also saw the entry of institutional investors and mainstream financial institutions into the cryptocurrency space. Major financial firms began exploring their own blockchain projects and cryptocurrencies, signaling a shift in the industry's legitimacy and potential for widespread adoption (Bouoiyour and Selmi 2019). Additionally, regulatory frameworks have started to evolve, aiming to address the challenges and opportunities presented by the growing cryptocurrency market. This increased institutional interest and regulatory development further solidified cryptocurrencies' position as a significant financial innovation.

The debate continues over whether cryptocurrencies constitute a distinct asset class continues. Despite functioning as digital mediums of exchange, cryptocurrencies lack the stability and governmental support that traditional currencies such as the US Dollar (USD) or Euro (EUR) have, leading to significant market volatility (Ciaian et al. 2016; Cheah and Fry 2015). It is not unusual for cryptocurrency market values to fluctuate by 20–30% within a single day, with substantial gains or losses occurring over short periods. For example, the cryptocurrency market saw a decline of over 75% from its peak in December 2017 to October 2018 (CoinMarketCap, 2018 accessed on 4 May 2024). According to Yermack (2015), Bitcoin's scarcity and volatility are reasons it cannot be considered a"real"currency. Team (2023) data shows that Bitcoin is predominantly held as an investment rather than used for transactions, with long-term investors holding 6 million Bitcoin, compared to 5 million held by short-term speculators (Cunha and Murphy 2019). The introduction of stablecoins, designed to reduce price volatility by pegging their value to fiat currencies, is seen as an effort to address these issues (Moin et al. 2019).

Regulatory approaches to cryptocurrencies differ greatly as authorities strive to balance their potential advantages with the necessity for legislative oversight. Many cryptocurrency exchanges are obliged to follow Know Your Customer (KYC) and Anti-Money Laundering (AML) laws, although in many places the market is still mostly uncontrolled (Sompolinsky and Zohar 2015). This regulatory vacuum, combined with the speculative nature of cryptocurrencies and the absence of governmental or institutional backing, contributes to market volatility and susceptibility to manipulation. For example, Gandal et al. (2018) looked into price manipulation on the Mt. Gox exchange and showed how a single actor may drive up the price of Bitcoin dramatically. Similarly, Griffin and Shams (2018) demonstrated that the cryptocurrency exchange Bitfinex used Tether to manipulate Bitcoin prices on a large scale in 2017. Moreover, the introduction of decentralized exchanges (DEXs) has posed new regulatory challenges due to their inherent resistance to censorship and control (Werner et al. 2022).

Numerous scholars and financial analysts argue that the cryptocurrency market has characteristics of asset bubbles and compare it to past occurrences such as the 1999 DotCom boom and the Dutch Tulip Mania of 1637 (Sovbetov 2018; Phillips and Gorse 2017). These research's findings imply that the market has experienced several bubbles over its existence. The market behaviour mimics the euphoric and maniacal conditions of investment that Aliber et al. (2023) describes, including the broad adoption of revolutionary technologies and speculative buying motivated by projections of future price increases. Critics of cryptocurrencies often argue that they lack intrinsic value, as their worth is not derived from discounted future cash flows but rather from speculative resale value (Mai et al. 2018). Additionally, the development of initial coin offerings (ICOs) has contributed to speculative behavior, with many projects raising significant funds based on future promises rather than tangible products (Howell et al. 2020).

The potential of cryptocurrencies serving as a hedge against political and financial instability is a significant research area of interest. Bitcoin has been demonstrated in research by Briere et al. (2015); Wang et al. (2019); Bouri et al. (2017) to be a useful hedge against global uncertainty and a portfolio diversifier across a variety of indices, currencies, and commodities. Additionally, Urquhart (2017) also discovered evidence of price clustering in Bitcoin, while Katsiampa (2017) looked into the volatility and predictability of cryptocurrencies using Generalised Autoregressive Conditional Heteroskedasticity (GARCH) modelling techniques. Furthermore, the rise of decentralized finance (DeFi) platforms has introduced new ways for investors to manage risk and hedge against market volatility through innovative financial instruments (Werner et al. 2022).

2.1.2 Forecasting and market efficiency

A sizable amount of research concentrating on cryptocurrency price forecasting has resulted from the growing interest in these currencies. Cryptocurrency markets are characterized by high volatility and unpredictability, which presents unique challenges compared to traditional financial markets. Numerous methods, ranging from statistical models to machine learning and deep learning, have been investigated by researchers in order predict cryptocurrency prices. Early attempts at forecasting cryptocurrency prices mostly made use of conventional time series analysis methods such Autoregressive Integrated Moving Average (ARIMA) models and GARCH models (Katsiampa 2017). While these models can capture linear dependencies and volatility clustering, they often fall short in accounting for the nonlinear and chaotic nature of cryptocurrency markets. Furthermore, research by Ciaian et al. (2018) has shown how the Bitcoin and other altcoin markets are interdependent, which contradicts the Efficient Market Hypothesis (EMH) that assumes market prices represent all available information (Fama 1970).

The Efficient Market Hypothesis (EMH) asserts that asset prices fully reflect all available information, making systematic prediction strategies ineffective. However, this hypothesis, particularly its strong form, has been challenged in cryptocurrency markets, which are characterized by irrational investor behavior, speculative trading, and information asymmetry (Urquhart 2017). Sentiment-driven forecasting assumes that market participants react inefficiently to social media signals, creating temporary price inefficiencies that predictive models can exploit. This aligns with behavioral finance theories, which highlight how investor sentiment, especially in speculative markets, leads to price deviations from fundamental values (Shiller 2003). The predictive success of sentiment-based Bi-LSTM models in this study supports the notion that cryptocurrency markets may not be fully efficient, underscoring the value of integrating alternative data sources like sentiment analysis into financial modeling.

Studies have moved more recently to machine learning methods, which provide more adaptability and the capacity to model intricate patterns. For instance, Patel et al. (2015) compared the performance of machine learning models such as Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Random Forests in predicting stock prices, concluding that these models significantly outperform traditional statistical methods. Similar findings have been reported in the context of cryptocurrency forecasting, where machine learning models have shown promising results in capturing intricate market dynamics (McNally et al. 2018). The introduction of cryptocurrency futures by the Chicago Board Options Exchange (CBOE) and the Chicago Mercantile Exchange (CME) has given market players additional tools, which has improved price discovery and market efficiency (Bouri et al. 2019).

2.1.3 Financial markets sentiment and predictability

Sentiment analysis has emerged as a powerful tool for predicting cryptocurrency price movements by leveraging the wealth of information available on social media platforms like Twitter (Haritha and Sahana 2023). Several studies have explored the predictive power of public Twitter sentiment for cryptocurrency prices. In Kraaijeveld and De Smedt (2020), the forecasting capability of Twitter sentiment on the returns of nine main cryptocurrencies, such as Bitcoin, Ethereum, and Litecoin, was extensively analysed. This research underscores the potential of Twitter sentiment as a valuable indicator for cryptocurrency price movements.

Sentiment research has been researched in great detail in financial forecasting, especially since social media platforms like Twitter have emerged and offer real-time insights into public sentiment. One of the foundational studies in this field was published by Bollen et al. (2011), which demonstrated the ability of Twitter mood metrics can predict stock market movements. Their results sparked an explosion of research looking at how sentiment on social media may be used to forecast financial markets. Sentiment analysis techniques can be broadly classified into rule-based methods and machine learning approaches. Rule-based methods, such as the use of lexicons, rely on predefined lists of words and their associated sentiment scores. One prominent example is the VADER (Valence Aware Dictionary and Sentiment Reasoner) model, which is particularly effective for analyzing social media text due to its ability to handle slang, emoticons, and contextual nuances. VADER assigns sentiment scores to words and phrases, allowing for a nuanced analysis of sentiment intensity and polarity. In a study by Hutto and Gilbert (2014b), VADER achieved high accuracy in sentiment analysis tasks compared to other lexicon-based approaches.

2.1.4 Twitter sentiment analysis

Machine learning approaches involve training models on labeled datasets to automatically classify text based on sentiment. Techniques such as Naive Bayes, SVM, and deep learning models like Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) have been widely used in sentiment analysis (Pang et al. 2008; Lee et al. 2017). In particular, deep learning models have demonstrated great ability in capturing the sequential and contextual aspects of textual data. Long Short-Term Memory (LSTM) networks, introduced by Hochreiter and Schmidhuber (1997), address the vanishing gradient problem of traditional RNNs through a gating mechanism that allows them to retain information over longer periods. LSTMs have been successfully applied to various forecasting tasks, including stock price prediction and cryptocurrency price forecasting (Fischer and Krauss

2018). In the context of cryptocurrency forecasting, McNally et al. (2018) reported that their LSTM model achieved a Mean Absolute Error (MAE) of 10.92 and a Mean Squared Error (MSE) of 223.34 in predicting Bitcoin prices. Their study highlighted the potential of LSTM networks in capturing the temporal dependencies in cryptocurrency price data.

Several research have looked at using deep learning models and sentiment analysis to predict the price of cryptocurrencies. Liew and Wang (2016) utilized sentiment analysis of Twitter data to predict Bitcoin prices, employing LSTM networks to capture the temporal dependencies in the sentiment data. They found that when sentiment analysis is included, forecasting accuracy is much higher than with models that only use historical price data. They reported an R-squared value of 0.78 for their LSTM model, highlighting its strong predictive capability. More recently, researchers have turned to deep learning approaches to improve the accuracy of cryptocurrency price prediction using Twitter sentiment. Wu et al. (2021) proposed a new forecasting framework for Bitcoin price prediction using Long Short-Term Memory (LSTM) networks. The authors incorporated various features, including historical prices, trading volume, and Twitter sentiment, to train their LSTM model and achieved promising results in terms of forecasting accuracy. In order to better understand how LSTM networks might be used to predict cryptocurrency prices, Ider and Lessmann (2022) concentrated on the influence of investor sentiment extracted from tweets, Reddit posts, and news articles. The author used BERT-based classifiers to classify the sentiment of textual data, and then an LSTM model was fed this information to predict future cryptocurrency returns. A further important research by Mittal and Goel (2012) looked into the predictive power of Twitter sentiment on Bitcoin prices. They analysed tweet sentiment using a lexicon-based method, and then they utilised LSTM networks to predict Bitcoin prices. Their model achieved a notable improvement in prediction accuracy, with an MAE of 6.84 and an MSE of 60.29. The study underscored the importance of sentiment analysis in capturing market dynamics and improving forecasting models. Despite the promising results, several challenges remain in the application of sentiment analysis and deep learning models to cryptocurrency forecasting. One major challenge is the inherent noise and short-text nature of social media data, which can lead to inaccuracies in sentiment classification. Enhancing the robustness of sentiment analysis models through domain-specific lexicons and advanced NLP techniques is an ongoing area of research (Hutto and Gilbert 2014b). Moreover, the high volatility and speculative nature of cryptocurrencies pose additional challenges for forecasting models. Developing models that can adapt to rapid market changes and account for external factors such as regulatory news and macroeconomic indicators is crucial for improving prediction accuracy (Duan et al. 2021).

The literature review highlights the growing interest in leveraging Twitter sentiment for cryptocurrency price prediction and the various deep learning approaches that have been employed to improve forecasting accuracy. However, there is still room for further research to explore the comparative performance of different deep learning architectures, such as LSTM, GRU, Bi-Directional LSTM,and Temporal Attention based-models in the context of cryptocurrency price forecasting using Twitter sentiment analysis.

3 Methodology

In this section, we outline the methodology used to integrate sentiment analysis with deep learning models for cryptocurrency price forecasting. Our approach leverages social media sentiment from Twitter (X) to identify sentiment trends related to Bitcoin investments and combines this information with historical Bitcoin price data to enhance predictive accuracy. This fusion of financial and sentiment data provides a more comprehensive framework for understanding and predicting Bitcoin price movements.

Our methothodology is structured into four key components:

1. Data Collection and Preprocessing - Aggregating and consolidating Bitcoin price data with Twitter sentiment data, ensuring consistency and removing irrelevant or noisy data points.
2. Sentiment Analysis - Processing tweet data using both lexicon-based (VADER) and transformer-based (XLM-RoBERTa) sentiment models to extract sentiment scores that capture market sentiment variations.
3. Feature Engineering - Enhancing the dataset by incorporating technical indicators such as Moving Average Convergence Divergence (MACD), Relative Strength Index (RSI), volatility measures, and trend predictions to enrich model inputs.
4. Deep Learning-Based Price Forecasting - Employing Long Short-Term Memory (LSTM), Bi-Directional LSTM (Bi-LSTM), Gated Recurrent Units (GRU), and Temporal Attention Models to predict Bitcoin price movements. Model performance is evaluated using key error metrics (MAPE, RMSE, MAE), and interpretability is improved using SHAP (SHapley Additive exPlanations) values to understand feature importance.

This structured approach ensures a robust analysis of the relationship between market sentiment and Bitcoin price fluctuations, providing insights into the impact of sentiment-driven trading behavior.

3.1 Data collection

This study aims to predict Bitcoin's closing prices by integrating financial data with sentiment analysis derived from social media discussions on Twitter (X). The data collection process is structured into two distinct phases, ensuring a comprehensive dataset that captures both market trends and sentiment analysis.

The first phase involves gathering Bitcoin-related tweets from X, focusing on posts that contain relevant financial terms and discussions about Bitcoin. These tweets serve as a proxy for market sentiment, reflecting public perceptions, speculative behavior, and reactions to news events. To extract sentiment signals from the text data, we employ two sentiment analysis techniques: VADER, a widely used lexicon-based model, and XLM-RoBERTa, a transformer-based model fine-tuned for sentiment classification. The latter provides a more nuanced understanding of sentiment by leveraging deep learning representations. The collected tweets undergo preprocessing steps, including removal of duplicates, filtering of irrelevant content, and text normalization to ensure high-quality input for sentiment analysis.

The second phase focuses on acquiring historical Bitcoin price data, which is sourced from CoinMarketCap.com. This financial dataset includes daily closing prices, trading volume, and key technical indicators such as Moving Average Convergence Divergence (MACD), Relative Strength Index (RSI), and volatility measures. By incorporating these financial metrics alongside sentiment scores, the study aims to enhance predictive accuracy and identify patterns linking investor sentiment to price movements.

A detailed overview of the data collection methodology is illustrated in Fig. 1, highlighting the integration of both social media and financial data for enhanced prediction accuracy.

3.1.1 Data from Twitter

Tweets were sourced from an open-access dataset available on Suresh (2023), featuring real-time cryptocurrency tweets collected from February 5, 2021, to March 5, 2023. The collected Twitter data from Kaggle was obtained using the Tweepy library to access the Twitter API. First, authentication credentials (user key, user secret, access token, and access token secret) from the Twitter developer account were required. These credentials were passed to Tweepy to establish an API connection. The search criteria were defined using keywords and hashtags related to Bitcoin, such as

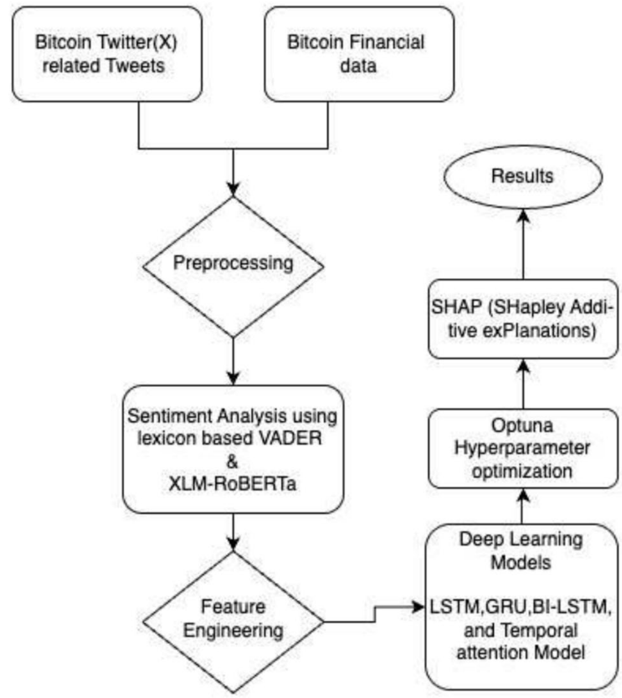

Fig. 1 An overall synopsis of the methodology's several stages

Table 1 The number of tweets prior to pre-processing

Cryptocurrency	Total number of tweets collected
Bitcoin (BTC)	48,674,883

("bitcoin","bitcoins","Bitcoin","Bitcoins","BTC","XBT", and"satoshi") or any hashtags of bitcoin's ticker symbols ("#bitcoin","#btc","$BTC","#XBT"). For each tweet, information such as tweet ID, content, username, follower count, retweet count, like count, and creation date was extracted.

The daily count of tweets related to BTC collected over the specified period is illustrated in Table 1. The distribution of Bitcoin-related tweets collected over the study period is illustrated in Fig. 2. A total of 48,674,883 tweets were gathered during this period, demonstrating substantial and sustained engagement with Bitcoin. The monthly distribution indicates variations in tweet volume, with notable peaks in April 2021, July 2021, and March 2022. These fluctuations reflect periods of increased discourse surrounding Bitcoin, which may be attributed to significant market developments. The dataset captures sentiment dynamics over an extended timeframe, allowing for a comprehensive analysis of social media sentiment and its potential relationship with Bitcoin price movements. As detailed in Fig. 1, the collected tweets underwent multiple pre-processing steps to ensure higher data quality and consistency. These included filtering out irrelevant metadata (e.g., user profile details), removing

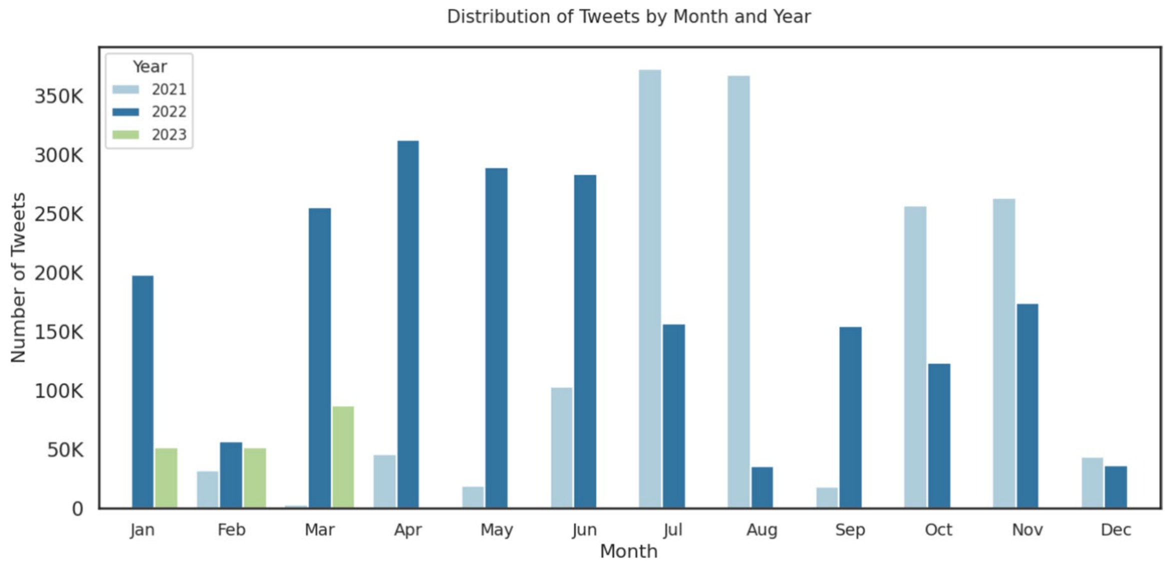

Fig. 2 Monthly distribution of Bitcoin-related tweets collected between February 5, 2021, and March 5, 2023. The chart illustrates tweet volume variations over time, capturing trends in social media discussions related to Bitcoin across different years

spammy or incomplete tweets, and normalizing the textual content (such as converting to lowercase and stripping URLs/hashtags). Additionally, duplicates and missing values were addressed, leaving only clean and meaningful text data for sentiment analysis.

3.1.2 Financial data

This study utilizes historical Bitcoin (BTC) price data obtained from the Yahoo Finance API, covering the period from February 5, 2021, to March 5, 2023. Yahoo Finance was selected as the data provider due to its ability to aggregate prices from multiple cryptocurrency exchanges, offering a comprehensive and unbiased representation of market trends. The dataset is updated daily and consists of five key attributes: high, low, open, close, and adjusted close. The high and low values represent the highest and lowest BTC prices recorded each day, while the open price refers to the first recorded transaction price at the start of the trading day. The close price, which is the primary focus of this study, represents the final recorded BTC price at the end of the trading day. Additionally, the adjusted close accounts for any market-adjusted corrections, such as dividends or stock splits, ensuring consistency in historical comparisons.

To explore the relationship between Bitcoin price fluctuations and sentiment-driven trading behavior, this study primarily focuses on daily closing prices. This choice is motivated by the need to align sentiment analysis with overall market trends, as daily price movements are more reflective of aggregated social media sentiment than high-frequency intraday fluctuations. As illustrated in Fig. 3, the blue line represents BTC's daily closing prices, showcasing the volatility of Bitcoin over the study period. The data highlights significant price fluctuations, with notable peaks above $60,000 in 2021, followed by a downward trend through 2022, and a period of relative stabilization in early 2023. This dataset serves as the foundation for the forecasting models, where sentiment analysis from Twitter is integrated with historical price movements to improve prediction accuracy.

3.1.3 Data pre-processing techniques and feature selection

Twitter data is naturally unstructured and frequently noisy, necessitating substantial pre-processing to make it suitable for sentiment analysis. Research has shown that unprocessed social media text can significantly degrade the accuracy of sentiment classifiers due to the presence of informal language, excessive punctuation, emojis, and special characters (Symeonidis et al. 2018; Zhao and Gui 2017). To ensure high-quality input data, we implemented a robust pre-processing pipeline consisting of 12 techniques, alongside additional domain-specific modifications aimed at preserving meaningful information while filtering out noise. These techniques were applied uniformly to both lexicon-based (VADER) and transformer-based (DeBERTa) sentiment analysis models.

We started with tokenization and normalization, which involved eliminating URLs, user mentions (@username), and extra whitespace. Given that URLs do not contribute to textual sentiment and user mentions introduce unnecessary noise, they were entirely removed. Retweets, identified by the"RT"prefix at the beginning of tweets, were filtered out to prevent duplicate sentiment signals. Additionally, tweets

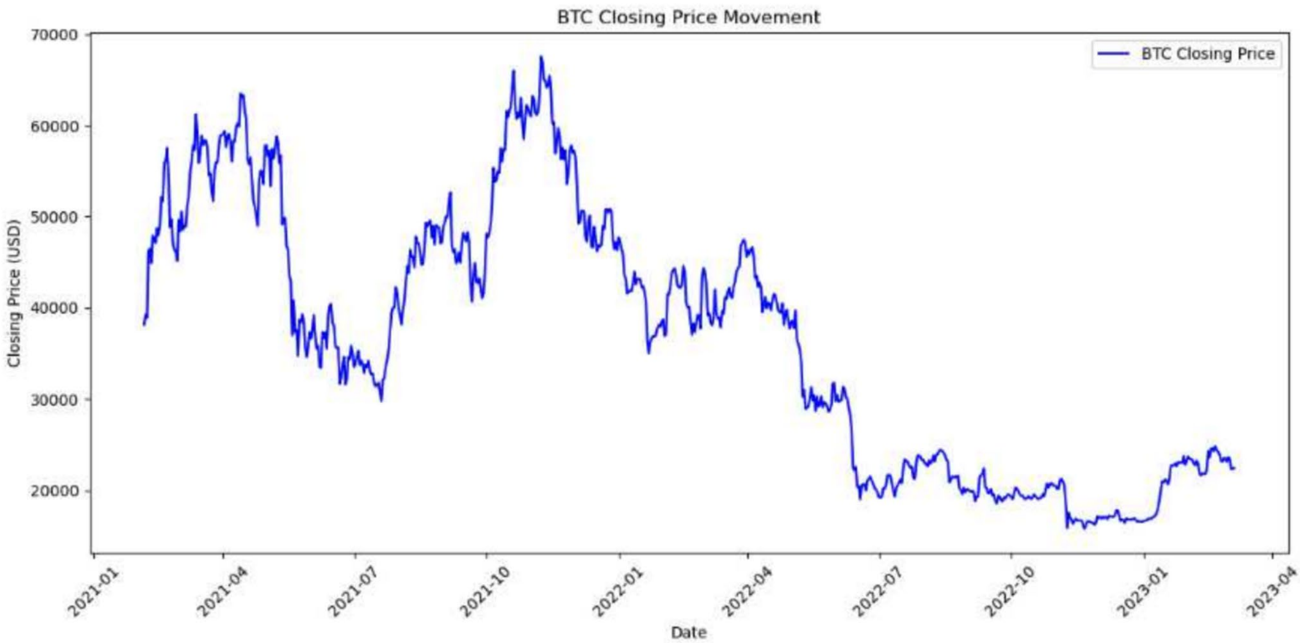

Fig. 3 Price movement of Bitcoin from the 25 th of February 2023 year to 5 th of March 2023 year

containing fewer than four tokens were excluded, ensuring that only sufficiently informative textual data remained. Research suggests that excessively short tweets often lack contextual sentiment, leading to unreliable sentiment scores (Hutto and Gilbert 2014a).

To preserve the sentimental value of hashtags, we devised a novel approach where hashtag tokens were evaluated against the NLTK Reuters English dictionary. If a hashtag was found in the dictionary, the"#"symbol was removed, treating it as a standard word; otherwise, the entire hashtag was omitted. This technique allowed the retention of semantically meaningful words while preventing non-standard abbreviations from distorting sentiment analysis. For instance, the tweet"Bitcoin is about to #surge! #buynow before it #hits $50000 tomorrow #btc #cryptocurrency"was processed into"Bitcoin is about to surge! Buy now before it hits $50000 tomorrow."This ensured that critical sentiment indicators embedded in hashtags were preserved while removing non-standard terminology (Zimbra et al. 2018).

Further refinements were made to expand contractions and normalize abbreviations, which are common in informal social media text. Examples include expanding"You're"to"You are","They've"to"They have", and"Didn't"to"Did not". Ticker symbols (e.g.,"*BTC*","ETH")-while useful in initial tweet filtering-were deemed extraneous for sentiment analysis and subsequently removed. Numbers embedded within words (e.g.,"1 st","2022","2nd") were similarly excluded as they did not contribute meaningful sentiment (Krouska et al. 2016). Cryptocurrency discourse is filled with domain-specific jargon, slang, and acronyms, which standard sentiment models often misinterpret. To mitigate this, we constructed a specialized cryptocurrency lexicon by aggregating terminology from various sources, including cryptocurrency forums, news articles, and online glossaries. Slang terms such as"HODL","FOMO","rekt", and"ATH"were manually mapped to their corresponding sentiment scores. Acronyms (e.g.,"LMAO","FUD","BTD") were expanded to their full meanings to ensure proper contextual understanding. Additionally, elongated words (e.g.,"Yesssssssss"→"Yessss","Aaaa ammmaaaazzzing"→"Amaaaazing") were truncated to three-character repetition limits, a technique known to reduce noise while retaining expressive sentiment cues (Effrosynidis et al. 2018). Punctuation and excessive special characters were eliminated, as they often introduce inconsistencies in tokenization. However, emojis and emoticons were preserved, given their proven effectiveness as sentiment indicators in lexicon-based and machine learning models (Hutto and Gilbert 2014a). Stopwords (e.g.,"I","them","it","this") were removed using an NLTK-customized English stopword list, ensuring that only sentiment-carrying words remained.

Handling of spam and promotional content was another essential step in refining the dataset. Tweets that contained spam-indicating keywords such as"Giveaway,""Airdrop," "Cashback,""NFT,""free money"were removed to prevent biased sentiment distributions. Similarly, user engagement-based filtering was applied, where tweets from accounts with fewer than 10 followers were flagged as potential bot-generated content and discarded (Symeonidis et al. 2018). Additionally, tweets with excessive hashtags were excluded to mitigate artificially inflated sentiment trends often introduced by promotional campaigns. To ensure temporal alignment between sentiment data and financial data, all timestamps were converted to UTC-1, addressing inconsistencies caused by Twitter's localized timestamp system. This standardization ensured that sentiment scores could be directly correlated with Bitcoin price movements on the same trading day.

Following the implementation of these rigorous preprocessing measures, the dataset was refined to 3,786,863 tweets, significantly improving the reliability of subsequent sentiment analysis and deep learning-based forecasting models. The reduction from over 48 million to approximately 3.7 million tweets is the cumulative result of the rigorous data cleaning steps outlined above. These procedures-including duplicate removal, bot filtering, language and lexicon normalization, spam exclusion, and short-text filtering-were essential to eliminate noise, ensure linguistic quality, and retain sentiment-relevant content. Sentiment classification using VADER and RoBERTa was applied only after preprocessing, as the unfiltered dataset was too noisy for meaningful sentiment extraction. Although this reduced the overall volume, the resulting dataset preserves a diverse sentiment distribution and remains representative of authentic Twitter discourse related to Bitcoin. This filtered corpus provides a more stable and interpretable basis for sentiment modeling and price forecasting. Table 2 illustrates the step-by-step transformations applied to a sample tweet, demonstrating how different techniques incrementally cleaned and structured the dataset.

3.1.4 Sentiment analysis using VADER

Sentiment polarity was computed using the Valence Aware Dictionary and Sentiment Reasoner (VADER), a lexicon-based algorithm designed specifically for analyzing informal text and social media, including cryptocurrency-related discussions (Hutto and Gilbert 2014b). VADER utilizes a human-curated lexicon (9,000 words) with predefined sentiment weights (ranging from negative to positive), accounting for capitalization, punctuation, modifiers, and emoticons common in social media discourse (Raheman et al. 2022; Pano and Kashef 2020). This method efficiently classifies sentiment without needing extensive labeled training data. Given the limitations of the standard VADER lexicon for domain-specific financial terms, particularly prevalent in cryptocurrency communities, we augmented VADER's vocabulary by incorporating specialized terms from the Loughran and McDonald financial corpus (Loughran and McDonald 2011) and additional cryptocurrency jargon (e.g., "HODL," "FOMO," "ATH," "rekt," "bullish," "bearish"). Prior research indicates that updating VADER with financial and cryptocurrency-specific terms notably enhances predictive accuracy in sentiment-driven cryptocurrency analyses

Table 2 An illustration of how pre-processing procedures are used

Processing approaches	Results
0. Initial Tweet	RT @crypto https://twitter.com/crypto/status/1234567890123456789 Bitcoin is about to surge! Buy now before it hits $50000 tomorrow #INVEST #CRYPTO #BTC #BLOCKCHAIN
1. If"RT"is present, eliminate it	@crypto https://twitter.com/crypto/status/1234567890123456789 Bitcoin is about to surge! Buy now before it hits $50000 tomorrow #INVEST #CRYPTO #BTC #BLOCKCHAIN
2. Remove URLs, extra spaces, and mentions	Bitcoin is about to surge! Buy now before it hits $50000 tomorrow #INVEST #CRYPTO #BTC #BLOCKCHAIN
3. Shorten repeated characters	Bitcoin is about to surge! Buy now before it hits $50000 tomorrow #INVEST #CRYPTO #BTC #BLOCKCHAIN
4. Convert to lowercase	bitcoin is about to surge! buy now before it hits $50000 tomorrow #invest #crypto #btc #blockchain
5. Discard short tweets	bitcoin is about to surge! buy now before it hits $50000 tomorrow #invest #crypto #btc #blockchain
6. Exclude non-Reuters corpus hashtags	bitcoin is about to surge! buy now before it hits $50000 tomorrow #invest #crypto #btc #blockchain
7. Expand short forms	bitcoin is about to surge! buy now before it hits $50000 tomorrow #invest #crypto #btc #blockchain
8. Process slang and acronyms	bitcoin is about to surge! buy now before it hits $50000 tomorrow #invest #crypto #btc #blockchain
9. Omit ticker symbols	bitcoin is about to surge! buy now before it hits $50000 tomorrow #invest #crypto #btc #blockchain
10. Remove words with numbers	bitcoin is about to surge! buy now before it hits $50000 tomorrow #invest #crypto #btc #blockchain
11. Delete common stop words	bitcoin surge! buy hits $ tomorrow invest crypto btc blockchain

(Garcia and Schweitzer 2015). Sentiment scores produced by VADER range from −1 (highly negative) to +1 (highly positive), classifying tweets as positive (> 0.05), neutral (−0.05 to 0.05), or negative (< −0.05). Daily sentiment scores were averaged to create a time-series sentiment measure. Additionally, to quantify opinion divergence and emotional intensity, we introduced a polarization score calculated via the geometric mean of average positive and negative sentiments within each timeframe, following methodologies from previous studies (Garcia and Schweitzer 2015).

The sentiment classification approach is illustrated in Fig. 4. Despite VADER's efficiency and interpretability, it has notable limitations. Lexicon-based models inherently struggle with rapidly evolving slang and context-specific terminology frequently used in cryptocurrency discourse (e.g., "to the moon," "diamond hands"), requiring continuous lexicon updates (Raheman et al. 2022). Furthermore, VADER's heuristic-driven approach lacks deeper contextual awareness, limiting its ability to interpret sarcasm, irony, or nuanced multi-sentence tweets accurately. Tweets containing mixed sentiments may often be classified as neutral due to the linear aggregation of individual valence scores.

Nevertheless, with appropriate domain-specific enhancements, VADER remains valuable for sentiment-driven financial forecasting, offering transparency, computational efficiency, and interpretability crucial for cryptocurrency sentiment analysis.

The distribution of sentiment scores derived from the VADER sentiment analysis provides crucial insights into the overall market sentiment regarding BTC over the studied period. Figure 5 presents three key visualizations: (a) the categorical distribution of sentiment labels (positive, neutral, and negative), (b) the density distribution of VADER compound sentiment scores, and (c) the temporal trend of VADER sentiment scores. These analyses illustrate

Fig. 4 Polarity Classification

Value of Polarity	Sentiment
> 0.05	Positive
0	Neutral
< 0.05	Negative

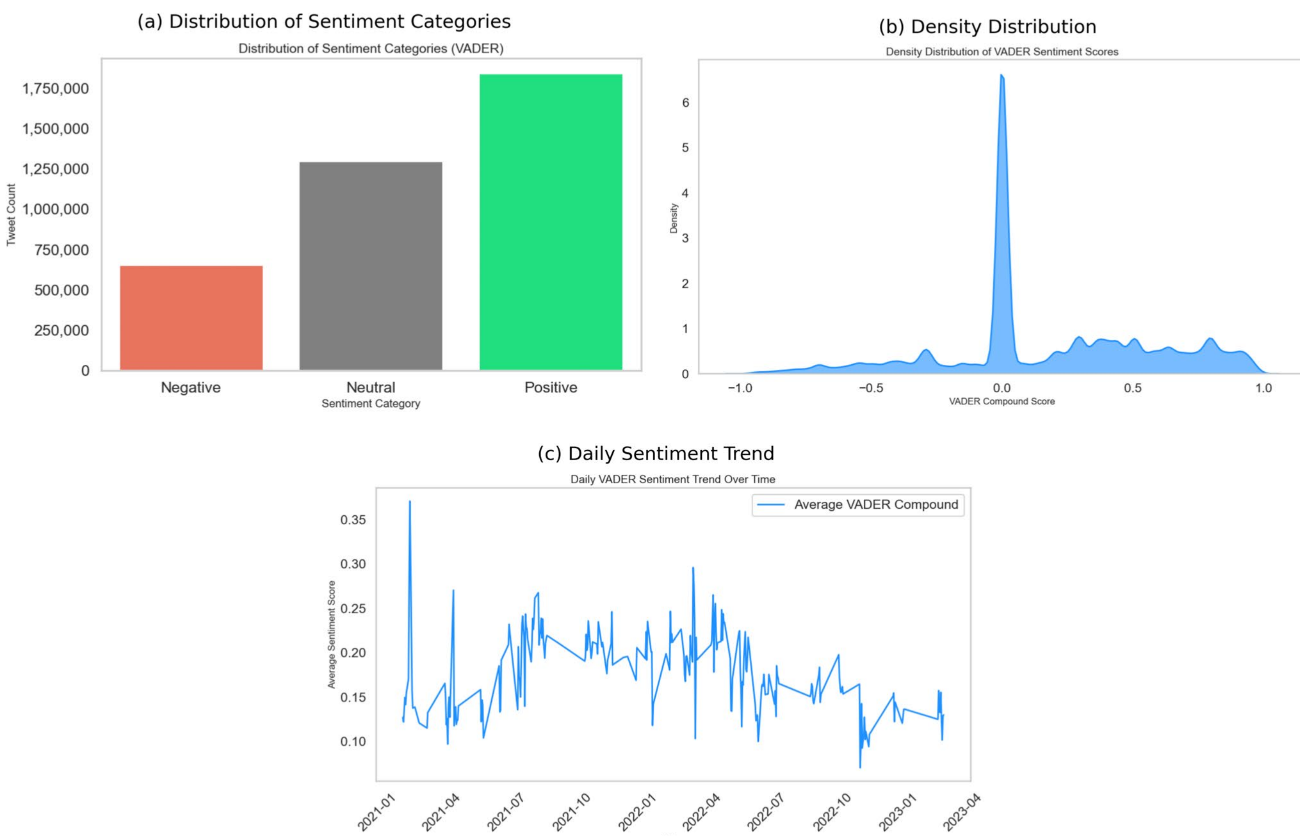

Fig. 5 VADER Sentiment Analysis: (**a**) Distribution of Sentiment Categories, (**b**) Sentiment Score Distribution, (**c**) Daily Sentiment Trend

the sentiment landscape of cryptocurrency discussions on Twitter(X) and how these sentiments fluctuate over time.

As shown in Fig. 5a, the sentiment category distribution indicates that positive sentiment dominates, followed by neutral and negative sentiments. This suggests that cryptocurrency discussions tend to be optimistic, aligning with the general enthusiasm of market participants. However, the significant proportion of neutral tweets reflects a mix of factual news and discussions that do not carry strong emotional tones. The temporal trend of VADER sentiment scores, illustrated in Fig. 5c, highlights fluctuations corresponding to major market events, with sentiment peaks aligning with bullish market conditions and sentiment dips occurring during downturns. Additionally, the density distribution of sentiment scores (Fig. 5b) further reinforces this trend, showing a strong concentration around neutral sentiment, with smaller peaks for both positive and negative extremes. This underscores the polarized nature of cryptocurrency discussions, where sentiment can quickly shift based on market developments. These insights confirm that social media sentiment plays a vital role in shaping market perception, making it a valuable component for forecasting Bitcoin price movements.

3.1.5 Sentiment analysis using transformer-based model (RoBERTa)

In contrast to VADER's lexicon-based approach, RoBERTa employs a transformer architecture that captures contextual nuances in sentiment. RoBERTa, a variant of BERT optimized for sentiment classification, has been pre-trained on large text corpora and fine-tuned on domain-specific data. The specific model utilized in this study, Twitter-XLM-RoBERTa-Bitcoin-Sentiment, extends a multilingual RoBERTa (XLM-RoBERTa) trained on approximately 58 million tweets SVA Labs. This model was further fine-tuned on Bitcoin-related tweets to enhance its understanding of sentiment within cryptocurrency discussions. Unlike lexicon methods, which rely on predefined word lists, RoBERTa encodes tweets into high-dimensional embeddings, allowing it to discern sentiment based on word relationships and context.

The transformer-based model's ability to capture intricate linguistic patterns makes it particularly effective in crypto-related discussions, where sentiment is often expressed through jargon, emojis, and evolving slang. For example, RoBERTa can distinguish between nuanced expressions such as "Bitcoin is not bad" and "not having Bitcoin is bad," recognizing that the latter expresses a stronger negative sentiment despite containing similar words. Additionally, since XLM-RoBERTa is inherently multilingual, it can process sentiment in tweets containing mixed languages, which is common in global cryptocurrency communities. Empirical research has demonstrated that fine-tuned transformer models achieve superior accuracy in sentiment classification tasks, with some studies reporting over 90% accuracy when classifying Bitcoin tweets, reinforcing their utility in financial forecasting models (Murugesapandian 2023).

Compared to VADER, RoBERTa operates under a fundamentally different paradigm. VADER provides a rule-based sentiment classification by assigning predefined polarity scores to words, whereas RoBERTa interprets sentiment dynamically by leveraging contextual embeddings. This allows RoBERTa to better handle sarcasm, idiomatic phrases, and informal crypto jargon. For instance, the phrase "Well, that's just great… Bitcoin crashed again" might be misinterpreted as positive by VADER due to the word "great," while RoBERTa, having been exposed to sarcastic patterns during training, would more accurately classify it as negative. Similarly, slang terms like "bagholder" (used negatively in crypto discussions) might be overlooked by VADER but correctly identified by RoBERTa through contextual inference.

Despite RoBERTa's advantages in contextual understanding, it requires extensive computational resources and labeled training data, whereas VADER is lightweight and interpretable. Studies have shown that carefully adapted lexicon-based approaches can still achieve strong correlations with market movements, particularly when supplemented with domain-specific vocabulary (Raheman et al. 2022). In practice, a combination of both methods can offer a more comprehensive sentiment analysis-VADER providing an interpretable baseline and RoBERTa delivering deeper contextual insights. The integration of both sentiment measures enhances robustness in financial sentiment modeling, ensuring that the strengths of lexicon-based and transformer-based approaches complement each other in cryptocurrency market forecasting. While lexicon-based approaches like VADER offer transparency and efficiency, recent research suggests that transformer-based models, such as RoBERTa, provide superior sentiment classification by leveraging deep contextual embeddings. Paul et al. (2024) propose a multi-tier deep learning framework for multimodal sentiment analysis, demonstrating that deep learning-based sentiment models outperform traditional lexicon-based approaches, particularly in finance and social media applications.

The sentiment analysis using RoBERTa presents a different perspective on cryptocurrency discussions compared to VADER. The sentiment category distribution, illustrated in Fig. 6, indicates that the majority of tweets are classified as neutral, significantly higher than both positive and negative sentiments. This suggests that RoBERTa captures a more cautious and context-aware sentiment distribution, likely due to its ability to understand nuanced expressions. Unlike VADER, which often assigns clear-cut

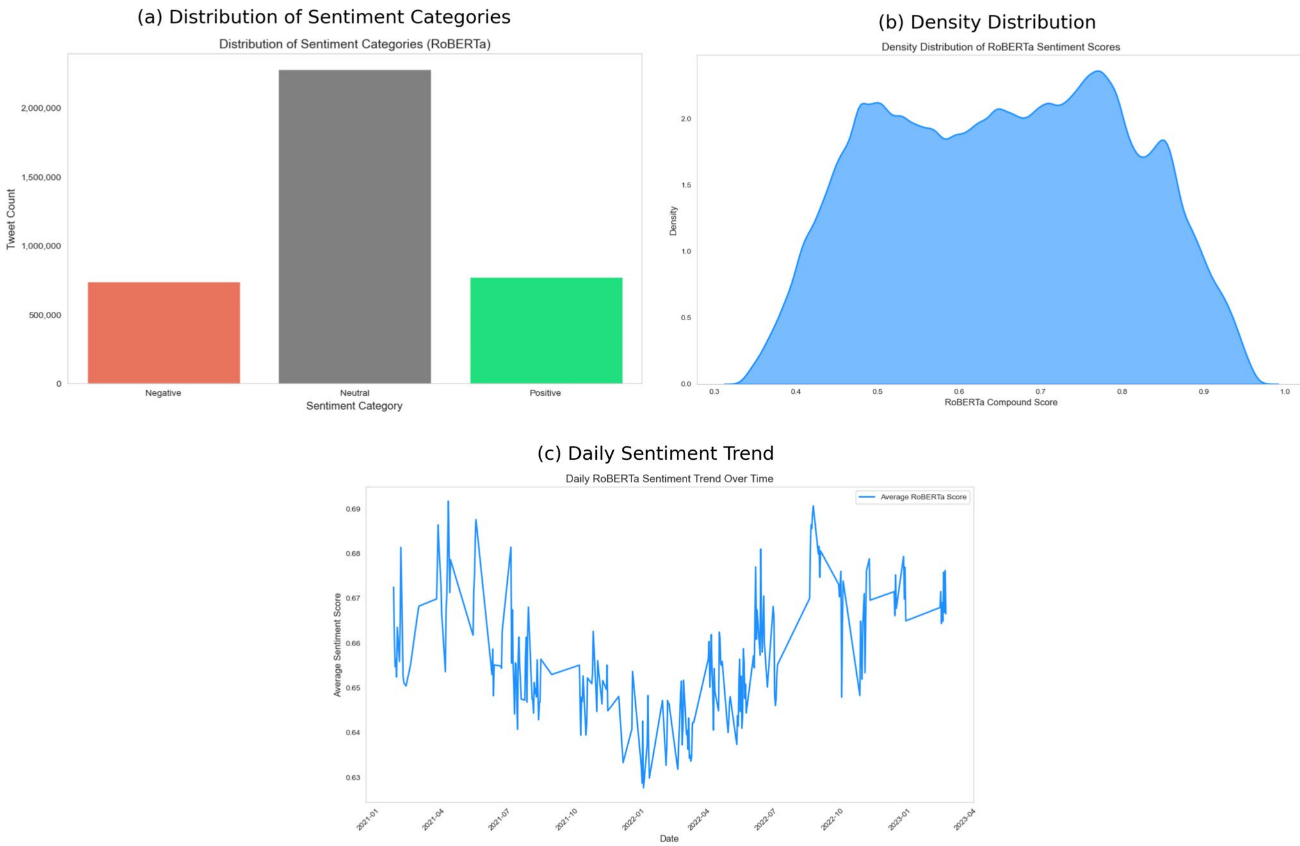

Fig. 6 RoBERTa Sentiment Analysis: (**a**) Distribution of Sentiment Categories, (**b**) Sentiment Score Distribution, (**c**) Daily Sentiment Trend (centered)

sentiment labels, RoBERTa's contextualized embeddings interpret many tweets as factual or mixed in tone rather than purely optimistic or pessimistic. The lower proportion of positive sentiment compared to VADER highlights the difference in methodology-RoBERTa, being transformer-based, relies on learned representations rather than predefined sentiment lexicons.

The temporal trend of RoBERTa sentiment scores, as shown in Fig. 6c, reveals fluctuations over time, with noticeable peaks and declines that align with key market movements. The sentiment trend shows periods of heightened optimism interspersed with downturns, reflecting the evolving sentiment in crypto-related discussions. The density distribution of RoBERTa scores, depicted in Fig. 6b, further emphasizes this pattern, showing a more continuous and less sharply peaked sentiment distribution compared to VADER. The absence of a strong concentration around zero suggests that RoBERTa is more effective at distinguishing varying levels of sentiment intensity rather than defaulting to neutrality. These insights indicate that RoBERTa's contextual learning provides a deeper understanding of sentiment trends, making it a valuable tool for analyzing cryptocurrency market discussions.

3.2 Feature engineering vectors

To capture both textual sentiment and quantitative market dynamics, we construct composite feature vectors for each time step t by incorporating sentiment analysis outputs from VADER and RoBERTa alongside cryptocurrency market features.

Sentiment Analysis Features VADER (Valence Aware Dictionary and sEntiment Reasoner) is a rule-based sentiment tool that provides sentiment scores calibrated for social media text (Hutto and Gilbert 2014a). For each relevant tweet in our dataset, VADER produces a four-dimensional sentiment output:

$$\mathbf{s}^{(\mathrm{VADER})} = \left[s^{(\mathrm{pos})}, s^{(\mathrm{neu})}, s^{(\mathrm{neg})}, s^{(\mathrm{compound})}\right], \tag{3.1}$$

where $s^{(\mathrm{pos})}$, $s^{(\mathrm{neu})}$, and $s^{(\mathrm{neg})}$ correspond to the positive, neutral, and negative sentiment proportions, and $s^{(\mathrm{compound})}$ is a normalized aggregation of the valence scores ranging from -1 (most negative) to $+1$ (most positive). We use $s^{(\mathrm{compound})}$ as a single quantitative measure of overall tweet sentiment.

In parallel, we leverage a RoBERTa-based language model pre-trained on social media data for sentiment classification SVA Labs. Specifically, we utilize the 'svalabs/twitter-xlm-roberta-bitcoin-sentiment' model, which is fine-tuned on Bitcoin-related tweets for improved sentiment classification. This model assigns softmax probabilities to three sentiment classes: *positive*, *neutral*, and *negative*:

$$\mathbf{s}^{(\text{RoBERTa})} = \left[P^{(\text{pos})}, P^{(\text{neu})}, P^{(\text{neg})}\right], \tag{3.2}$$

where $P^{(\text{pos})}$, $P^{(\text{neu})}$, and $P^{(\text{neg})}$ represent the softmax-normalized probabilities of a tweet belonging to each sentiment class. These probabilities are computed as:

$$P^{(c)} = \frac{e^{z_c}}{\sum_j e^{z_j}}, \tag{3.3}$$

where z_c is the logit output for class c (positive, neutral, or negative), and the softmax function ensures the sum of all class probabilities equals 1.

We define the sentiment composite score based on the highest-probability predicted sentiment:

$$s^{(\text{RoBERTa})} = \arg\max_c P^{(c)}, \tag{3.4}$$

where the probability of the selected sentiment class is used as the sentiment intensity score. The corresponding sentiment label is assigned based on the predicted class:

$$l^{(\text{RoBERTa})} = \begin{cases} 2, & \text{if } \arg\max(P^{(\text{pos})}, P^{(\text{neu})}, P^{(\text{neg})}) = P^{(\text{pos})}, \\ 1, & \text{if } \arg\max(P^{(\text{pos})}, P^{(\text{neu})}, P^{(\text{neg})}) = P^{(\text{neu})}, \\ -1, & \text{if } \arg\max(P^{(\text{pos})}, P^{(\text{neu})}, P^{(\text{neg})}) = P^{(\text{neg})}. \end{cases} \tag{3.5}$$

For each time interval t (e.g., a day), we aggregate sentiment scores:

$$V_t = \frac{1}{N_t}\sum_{i=1}^{N_t} s_i^{(\text{VADER})}, \quad R_t = \frac{1}{N_t}\sum_{i=1}^{N_t} s_i^{(\text{RoBERTa})}, \tag{3.6}$$

where N_t is the number of tweets in interval t.

3.3 Market feature engineering

For each time interval t, we extract the following market-based features, which capture price trends, volatility, and momentum:

Volatility: Estimated using rolling standard deviations over different time horizons:

$$\sigma_t^{(7)} = \text{std}(P_{t-6}^{(\text{close})}, \dots, P_t^{(\text{close})}), \quad \sigma_t^{(14)} = \text{std}(P_{t-13}^{(\text{close})}, \dots, P_t^{(\text{close})}). \tag{3.7}$$

Trend Direction: Defined as a binary movement indicator based on price changes:

$$T_t = \text{sign}(P_t^{(\text{close})} - P_{t-1}^{(\text{close})}). \tag{3.8}$$

Relative Strength Index (RSI): A momentum oscillator that quantifies recent price gains and losses:

$$\text{RSI}_t = 100 - \frac{100}{1 + RS_t}, \quad RS_t = \frac{\text{avg gain over 14 days}}{\text{avg loss over 14 days}}. \tag{3.9}$$

Moving Average Convergence Divergence (MACD): A trend-following momentum indicator computed as:

$$\text{MACD}_t = \text{EMA}_{12}(P_t^{(\text{close})}) - \text{EMA}_{26}(P_t^{(\text{close})}). \tag{3.10}$$

Lagged Closing Price: Incorporates past closing prices to capture short-term dependencies:

$$P_{t-1}^{(\text{close})}. \tag{3.11}$$

3.4 Sentiment-based features

To incorporate sentiment-driven dynamics, we engineer features reflecting sentiment momentum, volatility interactions, and rolling statistics:

Sentiment Momentum: Measures changes in sentiment volume (V_t) and sentiment ratio (R_t) over a 7-day period:

$$\Delta V_t = V_t - \frac{1}{7}\sum_{j=1}^{7} V_{t-j}, \quad \Delta R_t = R_t - \frac{1}{7}\sum_{j=1}^{7} R_{t-j}. \tag{3.12}$$

Sentiment-Volatility Interaction: Captures the impact of sentiment on market volatility:

$$S_t^{(\text{vol})} = V_t \cdot \sigma_t^{(14)}. \tag{3.13}$$

Rolling Sentiment Statistics: Aggregates sentiment scores over a rolling 7-day window:

$$\bar{V}_t^{(7)} = \frac{1}{7}\sum_{j=0}^{6} V_{t-j}, \quad \text{std}_t^{(7)} = \text{std}(V_{t-6}, \dots, V_t). \tag{3.14}$$

Lagged Sentiment Features: Incorporates past sentiment signals at various lookback periods:

$$V_{t-k}, \quad R_{t-k}, \quad k \in \{1, 2, 3, 5, 7\}. \tag{3.15}$$

Final Feature Vector By concatenating sentiment and market components, we form the final feature vector:

$$\mathbf{x}_t = \left[V_t, R_t, m_{t,1}, m_{t,2}, \dots, m_{t,k}\right]^\top. \tag{3.16}$$

This vector $\mathbf{x}_t$ represents the state of the system at time t in terms of both public sentiment and market conditions. By incorporating sentiment metrics derived from textual data (Twitter) together with numeric market data, we aim to capture a more holistic snapshot of factors influencing cryptocurrency price movements, as suggested by prior studies (Valencia et al. 2019). These feature vectors serve as inputs to our sequence models (LSTM, GRU, Bi-LSTM, and Temporal Attention Model), enabling the networks to learn from combined signals of crowd sentiment and market trends.

3.5 Normalization

Before feeding the data into neural network models, we apply normalization to ensure all features are on comparable scales. We employ the Min-Max scaling technique?, which transforms each feature into the range [0, 1]. For each feature j, we compute its minimum and maximum values across the training set:

$$x_{\min}^{(j)} = \min_t x_t^{(j)}, \quad x_{\max}^{(j)} = \max_t x_t^{(j)}. \tag{3.17}$$

Each raw value $x_t^{(j)}$ is then normalized as:

$$x_t^{(j),\text{scaled}} = \frac{x_t^{(j)} - x_{\min}^{(j)}}{x_{\max}^{(j)} - x_{\min}^{(j)}}, \quad x_t^{(j),\text{scaled}} \in [0, 1]. \tag{3.18}$$

This transformation ensures that all features lie within the unit interval [0, 1], preventing features with larger numerical ranges from dominating the neural network's learning process.

3.6 Sequence preparation and train-test data split

To train sequence-based models such as LSTM, GRU, Bi-LSTM, and Temporal Attention Models, we structure the dataset into rolling time windows of fixed length. For each time step t, we construct an input sequence containing past k time steps of feature vectors, where k represents the sequence length:

$$\mathbf{X}t = [\mathbf{x}t - k, \mathbf{x}t - k + 1, \dots, \mathbf{x}t - 1], \tag{3.19}$$

where each feature vector $\mathbf{x}_{t-i}$ consists of sentiment-based features (VADER and RoBERTa scores) and market-based indicators. The target variable is the closing price at time t:

$$y_t = P_t^{(\text{close})}. \tag{3.20}$$

Following standard machine learning practices, we split the dataset into training and validation sets, ensuring that validation data is never seen during training. We allocate 80% of the sequences to training and 20% to validation:

$$\mathcal{D}_{\text{train}} = (\mathbf{X}t, y_t) | t \le 0.8N, \quad \mathcal{D}\text{val} = (\mathbf{X}_t, y_t) | t > 0.8N, \tag{3.21}$$

where N is the total number of available sequences.

To ensure temporal integrity and avoid look-ahead bias, we adopted a time series-aware data splitting strategy. The sequences were divided chronologically, with the first 80% used for training and the remaining 20% reserved for validation. This approach preserves the natural order of observations, ensuring that the models are always evaluated on future data relative to the training window and closely mimicking real-world forecasting conditions.

3.7 Hyperparameter tuning

Hyperparameter tuning is crucial for optimizing model performance. We employ Optuna, a Bayesian optimization framework, to search for the optimal hyperparameters of the LSTM, GRU, Bi-LSTM, and Temporal Attention Models (Table 3). The following hyperparameters are optimized:

For the Temporal Attention Model (TAM), we optimized the number of attention heads (*num_heads* = 16) and feed-forward dimension (*ff_dim* = 256) to maximize feature extraction efficiency. Unlike traditional recurrent models, TAM dynamically weighs past observations, reducing noise in volatile market conditions. The optimal dropout rate (0.19997) and learning rate (0.00021) indicate that the model benefits from regularization while maintaining a low learning rate to prevent overfitting.

The objective function to minimize is the validation loss, defined as:

$$\mathcal{L}(\theta) = \frac{1}{|\mathcal{D}\text{val}|} \sum (\mathbf{X}t, y_t) \in \mathcal{D}\text{val}\left(y_t - \hat{y}_t\right)^2, \tag{3.22}$$

where $\hat{y}_t$ is the predicted closing price at time t. The optimization runs for multiple trials, selecting the best set of hyperparameters that minimizes $\mathcal{L}(\theta)$.

The Temporal Attention Model (TAM) employed in this study is not a novel architecture, but an adaptation of the self-attention mechanism introduced in the Transformer model by Vaswani et al. (2017). TAM is implemented as

Table 3 Model Hyperparameters

Model	Units/Heads	Dropout Rate	Learning rate	Batch size
LSTM	256 lstm_units	0.2791	0.00202	64
GRU	256 gru_units	0.2541	0.00165	32
Bi-LSTM	64 lstm_units	0.1602	0.00111	64
TAM	16 num_heads, 256 ff_dim	0.19997	0.00021	16

a standalone attention-based forecasting model comprising multi-head self-attention layers, position-wise feed-forward networks, and dropout regularization to enhance predictive robustness. While originally developed for natural language processing, this architecture has been adapted in our framework to capture temporal dependencies in financial time-series data. For benchmarking purposes, TAM is evaluated alongside recurrent models (LSTM, GRU, Bi-LSTM) using identical sentiment-based feature inputs. Although more advanced architectures like Informer (Zhou et al. 2021) further optimize attention mechanisms for long sequences, our implementation of TAM remains intentionally standard and interpretable for comparative analysis.

3.8 SHAP analysis for feature interpretation

To interpret the influence of sentiment and market-based features on price predictions, we utilize SHapley Additive exPlanations (SHAP). SHAP assigns an importance value ϕ_j to each feature x_j, quantifying its contribution to the prediction:

$$\phi_j = \sum_{S \subseteq \mathcal{F} \setminus j} \frac{|S|!(|\mathcal{F}| - |S| - 1)!}{|\mathcal{F}|!} \left[f(S \cup j) - f(S) \right], \qquad (3.23)$$

where $\mathcal{F}$ is the set of all features, and $f(S)$ represents the model's prediction given subset S.

We apply SHAP separately to both VADER and RoBERTa sentiment-enhanced models to compare the significance of sentiment-based and market-based features in price prediction. The resulting feature importance rankings indicate whether sentiment signals provide predictive power beyond traditional financial indicators.

4 Results

This section details the outcomes of our experiments in predicting cryptocurrency prices using Twitter sentiment data. We analyzed the correlation between different types of sentiments and cryptocurrency closing prices, informing our model selection and feature engineering. The predictive performance of LSTM, GRU, Bi-LSTM, and Temporal Attention model (TAM) models was evaluated, revealing the superior accuracy of the Bi-LSTM model.

4.1 Feature types and correlation results

In this study, we initially explore the correlation between Twitter sentiments and cryptocurrency closing prices to understand the influence of social media sentiment on market behavior. This preliminary analysis is crucial for developing effective predictive models by identifying the potential predictive power of sentiment features.

To explore sentiment-driven market behavior, we assessed correlations between BTC closing prices and three sentiment types (positive, neutral, negative). Scatter plots with regression lines Fig. 7 illustrate these relationships:

- Positive Sentiment: A slight positive correlation is observed. The regression line suggests marginal price increases as positive sentiment rises, aligning with market theories where positive sentiment boosts investor confidence and buying activity. However, the significant scatter indicates substantial influence from other factors, limiting sentiment as a standalone predictor.
- Neutral Sentiment: A slight negative correlation emerges, which is counterintuitive. Higher neutral sentiment might signal market uncertainty or investor indecision, reducing market participation or cautious trading behavior. Significant scatter indicates other stronger influences on prices.

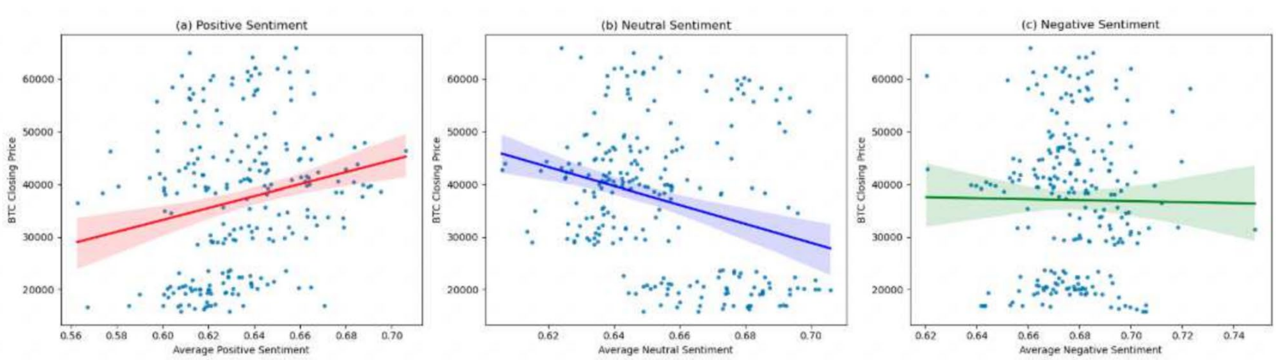

Fig. 7 Correlation between BTC closing prices and sentiment types: (**a**) Positive sentiment, (**b**) Neutral sentiment, and (**c**) Negative sentiment

– Negative Sentiment: Surprisingly, the analysis reveals a virtually negligible correlation between negative sentiment and downward price movements. The near-horizontal regression line suggests that negative sentiment alone is insufficient for predicting declines in asset prices. This finding may indicate that market participants perceive negative sentiment as transient noise or short-term volatility, thereby discounting its predictive value in forecasting price trends. Moreover, this aligns with behavioral finance theories, which suggest that investors may engage in contrarian trading strategies, where periods of extreme pessimism trigger buybacks as assets become undervalued. Research by Baker and Wurgler (2006); Tetlock (2007) supports this view, showing that sentiment-driven market fluctuations are often temporary and can lead to counterintuitive price movements as traders adjust their positions in response to perceived overreactions.

This correlation analysis highlights sentiment as a nuanced predictor rather than a standalone indicator, emphasizing its complementary role alongside other market factors. These insights guided our selection of predictive features and informed the development and training of LSTM, GRU, Bi-LSTM, and Temporal Attention Models, each trained with an 80-20% train-test split, enabling robust predictive performance evaluation.

4.2 Detailed results of LSTM, GRU, temporal attention model, and Bi-LSTM models

4.2.1 Performance evaluation of LSTM-based sentiment models

To assess the predictive power of sentiment-enhanced LSTM models, four key performance metrics were evaluated: Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Directional Accuracy. These metrics measure both absolute prediction errors and the model's ability to correctly identify market trends.

4.2.2 MAPE, RMSE, MAE, and directional accuracy results

See Table 4.

The results in Table 4 indicates that RoBERTa-based sentiment analysis significantly improves forecasting accuracy. The optimized model achieves a MAPE of 2.54%, compared to 4.83% for VADER, demonstrating that deep contextual sentiment models enhance price prediction.

Table 4 Performance metrics comparison for LSTM-based models

Metric	LSTM (RoBERTa)	LSTM (VADER)
MAPE (%)	2.54	4.83
RMSE	0.01	0.05
MAE	0.01	0.03
Directional Accuracy (%)	69.27	51.01

4.3 LSTM analysis of prediction accuracy: VADER vs. RoBERTa

The LSTM model enhanced by VADER sentiment demonstrates a general capability to track cryptocurrency price trends, with predicted prices aligning closely with actual prices during stable market conditions. However, its performance noticeably declines in periods of heightened volatility, exhibiting pronounced step-like patterns in predicted prices. These abrupt movements suggest that VADER's static lexicon-based sentiment scoring struggles to accurately reflect real-time, nuanced sentiment changes, particularly in speculative markets like cryptocurrencies. Prior research (Colianni et al. 2015; Fischer and Krauss 2018) confirms that lexicon-based methods like VADER have inherent limitations in capturing evolving language nuances, such as slang or sarcasm, leading to lower directional accuracy (51.01%) in predicting upward or downward price trends effectively.

In contrast, the RoBERTa-based LSTM model demonstrates significantly improved forecasting accuracy, as indicated by its smooth alignment with actual BTC prices, even during volatile periods. The model's predictions closely mirror market movements with minimal delays, highlighting RoBERTa's superior capability in capturing context-dependent sentiment shifts. The deep contextual embeddings provided by RoBERTa allow the model to adapt dynamically, substantially enhancing forecasting performance. This effectiveness is further validated by improved metrics such as reduced MAPE (2.54%) and increased directional accuracy (69.27%). Therefore, the RoBERTa-enhanced LSTM model emerges as a more robust and reliable method for real-time cryptocurrency price forecasting. A comparative analysis of the two models is illustrated in Fig. 8.

4.4 SHAP analysis for feature importance: VADER vs. RoBERTa

The SHAP (SHapley Additive exPlanations) analyses depicted in Fig. 9 illustrate the feature contributions for the LSTM models incorporating VADER and RoBERTa sentiment methods.

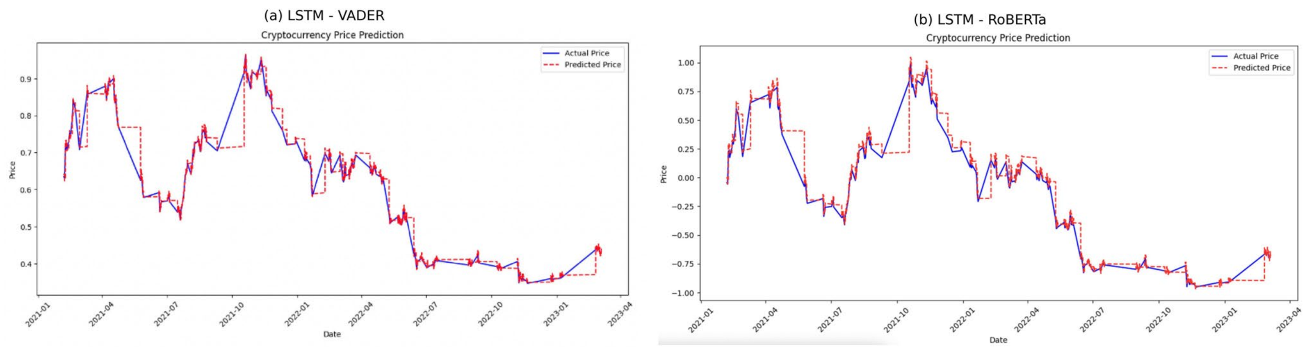

Fig. 8 Actual vs. Predicted BTC Prices using LSTM models: (**a**) VADER-based, (**b**) RoBERTa-based

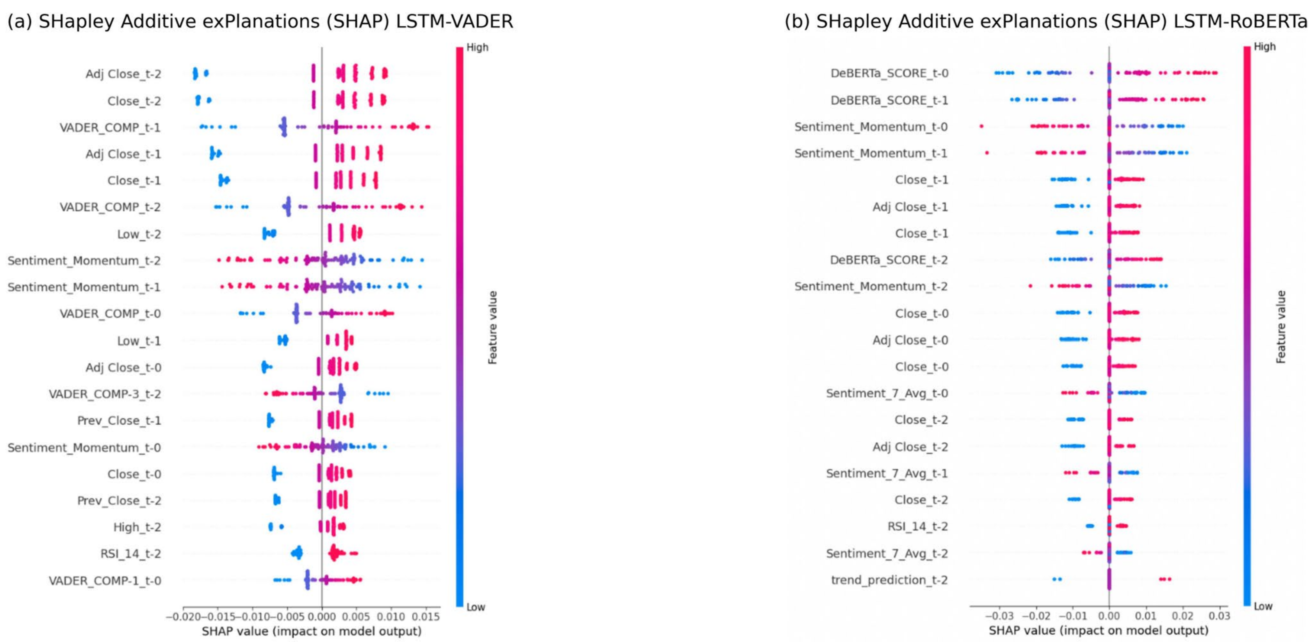

Fig. 9 SHAP analysis of feature importance for sentiment-enhanced LSTM models: (**a**) VADER-based features, (**b**) RoBERTa-based features

In the VADER-based LSTM model (Fig. 9a), features such as VADER_COMP$_{t-1}$, VADER_COMP$_{t-2}$, and Sentiment_Momentum$_{t-1}$ significantly drive model predictions, highlighting the critical role of recent sentiment and sentiment momentum. However, VADER's reliance on a static, lexicon-based sentiment scoring approach presents substantial limitations, as it struggles to capture the evolving, context-dependent language nuances common in speculative cryptocurrency markets, including slang, sarcasm, or emerging market terminologies (Medhat et al. 2014; Loughran and McDonald 2011). Consequently, these limitations likely contribute to the VADER-LSTM model's comparatively low directional accuracy (51.01%) (Garcia and Schweitzer 2015). Additionally, the inclusion of lagged features like Adj Close$_{t-1}$, Close$_{t-2}$, and technical indicators (e.g., RSI$_{14,t-2}$, High$_{t-2}$) aligns with the model's sequential nature, emphasizing the necessity of integrating historical price trends and sentiment measures for robust predictions (Fischer and Krauss 2018).

In contrast, the RoBERTa-based LSTM model (Fig. 9b) demonstrates strong feature influence from recent sentiment scores (DeBERTa_SCORE$_{t-0}$, DeBERTa_SCORE$_{t-1}$) and sentiment momentum (Sentiment_Momentum$_{t-0}$). This aligns with the expectation that recent sentiment scores and momentum are critical for forecasting in financial markets. RoBERTa's advanced contextual embeddings capture nuanced sentiment variations effectively, overcoming the shortcomings of static lexicon-based methods like VADER (Liu et al. 2019; Loughran and McDonald 2011). The presence of lagged financial indicators (Close$_{t-1}$, Adj Close$_{t-1}$) and technical indicators (RSI$_{14,t-2}$, trend_prediction$_{t-2}$) underscores the LSTM's capability to leverage historical data for temporal prediction, reinforcing that integrating advanced sentiment analysis with traditional financial indicators significantly enhances forecasting robustness (Zhang et al. 2018). The combined SHAP analysis confirms that context-aware sentiment analysis methodologies like RoBERTa offer substantial improvements over traditional lexicon-based methods, particularly in complex and speculative financial market environments.

The importance of explainability in sentiment-driven forecasting has been highlighted in recent literature. Explainability frameworks, such as SHAP, provide deeper insights into feature importance, ensuring transparency in model decision-making. Jothi Prakash and Arul Antran Vijay (2024) propose a multi-aspect framework for explainable sentiment analysis, emphasizing the necessity of interpretability when deploying sentiment-driven models in financial forecasting. This aligns with our use of SHAP to quantify the impact of sentiment signals on cryptocurrency price predictions.

4.5 Performance evaluation of GRU-based sentiment models

To assess the predictive capability of sentiment-enhanced GRU models, four key performance metrics were analyzed: Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Directional Accuracy. These metrics provide insights into both absolute prediction accuracy and the model's ability to correctly identify price movements.

Table 5 Performance Metrics Comparison for GRU-based Sentiment Models

Metric	GRU (RoBERTa)	GRU (VADER)
MAPE (%)	4.36	5.5
RMSE	0.02	0.08
MAE	0.06	0.05
Directional Accuracy (%)	62.3	50.02

4.5.1 MAPE, RMSE, MAE, and directional accuracy results

See Table 5.

4.6 GRU analysis of prediction accuracy: VADER vs. RoBERTa

The GRU models enhanced by VADER and RoBERTa sentiment analyses exhibit distinct performance characteristics in cryptocurrency price forecasting, as illustrated in Fig. 10. The RoBERTa-based GRU model in Table 5 achieves improved forecasting accuracy, evident from its lower MAPE (4.36%) compared to the VADER-based GRU model (5.5%), underscoring the effectiveness of transformer-based sentiment embeddings in financial forecasting.

The GRU model utilizing VADER sentiment analysis demonstrates a reasonable ability to capture macro-level trends during stable market conditions, as evidenced by its general alignment with actual prices. However, the model noticeably struggles during periods of high volatility, displaying abrupt, step-like corrections in predictions. These limitations stem from the inherent constraints of GRU's simplified gating mechanism, which may inadequately capture long-term dependencies crucial in financial forecasting (Cho et al. 2014a). Additionally, VADER's static lexicon-based sentiment scoring approach is limited in adaptability, rendering it less effective in capturing evolving and nuanced cryptocurrency market sentiment. Consequently, the GRU-VADER model exhibits delayed responses and inaccuracies during rapid price fluctuations, further reinforced by its higher RMSE (0.08) and low directional accuracy (50.02%). Prior studies (Nassirtoussi et al. 2015) indicate lexicon-based sentiment analysis methods' susceptibility to inconsistencies arising

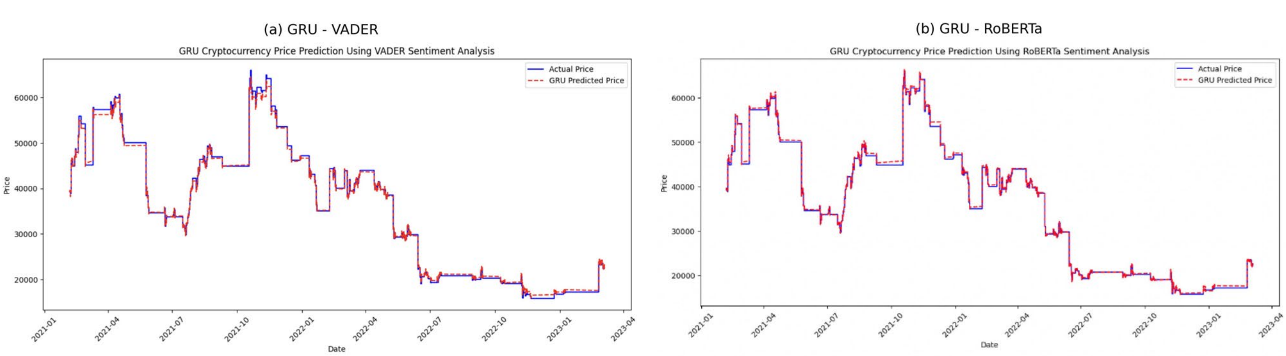

Fig. 10 Comparison of Actual vs. Predicted BTC Prices: (**a**) GRU-VADER Model, (**b**) GRU-RoBERTa Model

from dynamic contexts, emphasizing the need for more contextually adaptive approaches.

In contrast, the RoBERTa-based GRU model shows stronger alignment with actual market trends, achieving better directional prediction accuracy. Nonetheless, it shares similar limitations with the VADER-GRU model regarding abrupt step-like transitions during volatile periods, suggesting inherent architectural constraints in the GRU's handling of smooth, real-time price movements (Cho et al. 2014a). This behavior aligns with findings by Gao et al. (2021), suggesting that neither GRU nor LSTM consistently outperforms the other comprehensively, though other research (Xiao et al. 2024) supports LSTM's superiority in capturing long-term financial patterns due to better memory retention. RoBERTa's advanced sentiment embeddings enhance the model's predictive power through nuanced contextual understanding (Liu et al. 2019), yet the overall performance is still constrained by GRU's structural characteristics. These results highlight the potential benefits of exploring hybrid or alternative model architectures capable of effectively integrating advanced sentiment analysis with robust long-term dependency modeling.

4.7 GRU SHAP analysis: VADER vs. RoBERTa

The SHAP (SHapley Additive exPlanations) analyses illustrated in Fig. 11 provide insights into feature contributions for GRU models incorporating VADER and RoBERTa sentiment methods.

For the GRU model utilizing VADER sentiment analysis (Fig. 11a), historical price-related features (Close_{t-2}, Adj Close_{t-2}, Low_{t-2}, and Close_{t-1}) prominently influence predictions, suggesting that the GRU-VADER model relies more on past price movements rather than sentiment-based features. In comparison, sentiment-related features such as VADER sentiment scores (VADER_COMP_{t-0}, $t-1$, $t-2$) and sentiment momentum indicators ($\text{Sentiment\_Momentum}_{t-0}$, $t-1$, $t-2$) demonstrate considerably weaker impact as compared to historical price-related variables. This disparity highlights the limited predictive power of VADER's static lexicon-based sentiment analysis, suggesting it does not adequately capture the nuanced signals required for effective financial forecasting.

In contrast, the RoBERTa-based GRU model (Fig. 11b) exhibits significant contributions from both historical price features (Close_{t-2}, Close_{t-1}, Adj Close_{t-2}) and sentiment-derived features ($\text{DeBERTa\_SCORE}_{t-2}$, $\text{Sentiment\_Momentum}_{t-2}$, $\text{Sentiment\_7\_Avg}_{t-2}$). Notably, the higher SHAP values for RoBERTa-based sentiment features across multiple lags ($t-0$, $t-1$, $t-2$) indicate a stronger predictive influence compared to the VADER-based model. These findings underscore RoBERTa's superior capability in capturing contextually rich sentiment signals relevant to financial market movements. However, historical price trends continue to dominate the GRU's prediction process, emphasizing the model's tendency to prioritize short-term dependencies and potentially limiting its integration of longer-term sentiment trends. Overall, the SHAP analyses highlight the strengths of transformer-based sentiment

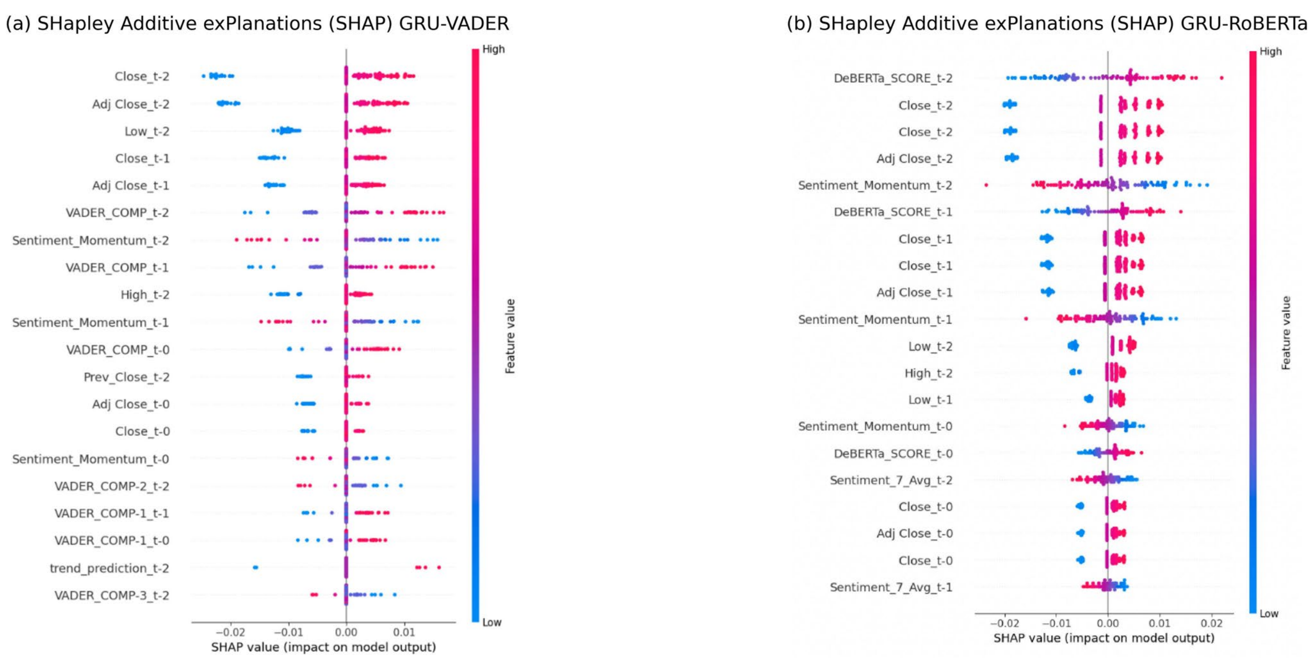

Fig. 11 SHAP Feature Importance for GRU models: (**a**) VADER-based features, (**b**) RoBERTa-based features

embeddings like RoBERTa over traditional lexicon-based methods such as VADER, while also indicating GRU's inherent preference for recent price history over longer-term sentiment signals.

4.8 Performance evaluation of Bi-LSTM based sentiment models

To assess the predictive capability of sentiment-enhanced Bi-LSTM models, four key performance metrics were analyzed: Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Directional Accuracy. These metrics provide insights into both absolute prediction accuracy and the model's ability to correctly identify price movements.

4.9 MAPE, RMSE, MAE, and directional accuracy results

See Table 6.

4.10 Bi-LSTM analysis of prediction accuracy: VADER vs. RoBERTa

The Bi-LSTM model demonstrates significant improvements over traditional LSTM and GRU architectures for sentiment-driven cryptocurrency price forecasting. By processing sequences bidirectionally, Bi-LSTM effectively integrates past dependencies and anticipates future trends, enhancing sentiment analysis integration and market trend detection. This capability is reflected in superior performance metrics as shown in Table 6, including lower MAPE (2.01% for RoBERTa, 3.90% for VADER), higher directional accuracy (79.5% for RoBERTa, 70.1% for VADER), lower RMSE (0.01 for RoBERTa, 0.05 for VADER), and reduced MAE (0.015 for RoBERTa, 0.025 for VADER). These results support findings by Zhang et al. (2018); Yu et al. (2019), and Fischer and Krauss (2018), emphasizing the effectiveness of bidirectional architectures in financial forecasting tasks, particularly those heavily influenced by sentiment.

Table 6 Performance Metrics Comparison for Bi-LSTM-Based Sentiment Models

Metric	Bi-LSTM (RoBERTa)	Bi-LSTM (VADER)
MAPE (%)	2.01	3.90
RMSE	0.01	0.05
MAE	0.015	0.025
Directional Accuracy (%)	79.5	70.1

Specifically, the VADER-based Bi-LSTM model, illustrated in Fig. 12a, outperforms both LSTM and GRU models by providing smoother and more stable predictions aligned closely with actual cryptocurrency prices. This improvement demonstrates Bi-LSTM's capability to effectively capture bidirectional sequence dependencies, significantly enhancing price trend recognition. Despite the inherent limitations of VADER's static lexicon-based sentiment scoring approach, Bi-LSTM achieves notable forecasting accuracy improvements, evidenced by a directional accuracy of 70.1% and reduced MAPE (3.90%), highlighting its potential for more reliable sentiment-driven forecasting.

The RoBERTa-based Bi-LSTM model, depicted in Fig. 12b, exhibits superior forecasting accuracy and an exceptional ability to align closely with actual market trends. Its predictions follow actual price movements precisely, demonstrating minimal lag, even during highly volatile periods. RoBERTa's advanced contextual sentiment embeddings, combined with Bi-LSTM's bidirectional sequence learning, enable the model to effectively capture nuanced sentiment dynamics and rapidly shifting market conditions. This capability is validated by notably higher directional accuracy (79.5%) and significantly lower MAPE (2.01%), confirming the robustness of the RoBERTa-Bi-LSTM combination for accurate, real-time cryptocurrency price predictions. Overall, these findings collectively demonstrate the

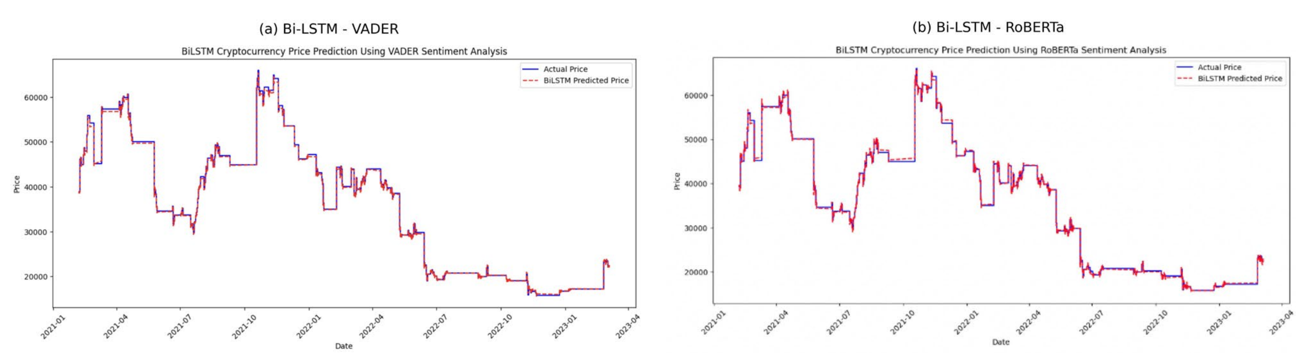

Fig. 12 Comparison of Actual vs. Predicted BTC Prices: (**a**) Bi-LSTM-VADER Model, (**b**) Bi-LSTM-RoBERTa Model

advantages of integrating advanced context-aware sentiment analysis methods with bidirectional architectures for financial forecasting, significantly outperforming traditional lexicon-based models.

4.11 Bi-LSTM SHAP analysis: VADER vs. RoBERTa

The SHAP analyses illustrated in Fig. 13 highlight the feature contributions for Bi-LSTM models employing VADER and RoBERTa sentiment analyses, respectively.

The Bi-LSTM model enhanced by VADER sentiment (Fig. 13a) predominantly relies on historical price features such as Adj Close_{t-2}, Close_{t-2}, and Low_{t-2}. While sentiment features including $\mathrm{VADER\_COMP}_{t-0}$, $t-1$, $t-2$, $t-3$, and sentiment momentum indicators ($\mathrm{Sentiment\_Momentum}_{t-1}$, $t-2$) exhibit some influence, their SHAP values remain relatively modest. This indicates the limited predictive power of VADER's lexicon-based sentiment analysis in effectively capturing nuanced financial market sentiment, aligning with previous findings about its inherent limitations. Despite these limitations, Bi-LSTM still improves the overall contribution of sentiment-based features compared to LSTM and GRU, demonstrating its ability to extract useful patterns even from less sophisticated sentiment models.

Conversely, the RoBERTa-enhanced Bi-LSTM model displays significant feature contributions from both sentiment-derived and historical price features. Features such as $\mathrm{DeBERTa\_SCORE}_{t-2}$, $\mathrm{DeBERTa\_SCORE}_{t-1}$, and $\mathrm{Sentiment\_Momentum}_{t-0}$ show prominently high SHAP values (Fig. 13b), underscoring RoBERTa's superior capability in capturing nuanced sentiment signals relevant to financial forecasting. Historical price features (Close_{t-2}, Adj Close_{t-2}, and technical indicators like $\mathrm{RSI}_{14,t-2}$) also significantly influence model predictions, demonstrating Bi-LSTM's robust integration of both sentiment and traditional market indicators.

These SHAP analyses confirm the effectiveness of bidirectional sequence learning provided by Bi-LSTM and highlight the enhanced predictive accuracy achievable when combined with transformer-based, contextually rich sentiment embeddings from RoBERTa, as opposed to traditional lexicon-based approaches such as VADER.

4.12 Performance evaluation of temporal attention model-based sentiment models

The Temporal Attention Model (TAM) is based on the Transformer architecture, which is designed to capture long-term dependencies in sequential data through self-attention mechanisms (Vaswani et al. 2017). Unlike recurrent architectures such as LSTM and GRU, which process data sequentially, Transformers analyze all time steps simultaneously, allowing them to extract global dependencies more efficiently. This characteristic makes them particularly well-suited for financial forecasting, where long-term trends and contextual relationships between market movements are crucial (Zerveas et al. 2021). The model architecture integrates multi-head self-attention, feed-forward layers, and dropout regularization to enhance predictive robustness. The motivation behind employing this

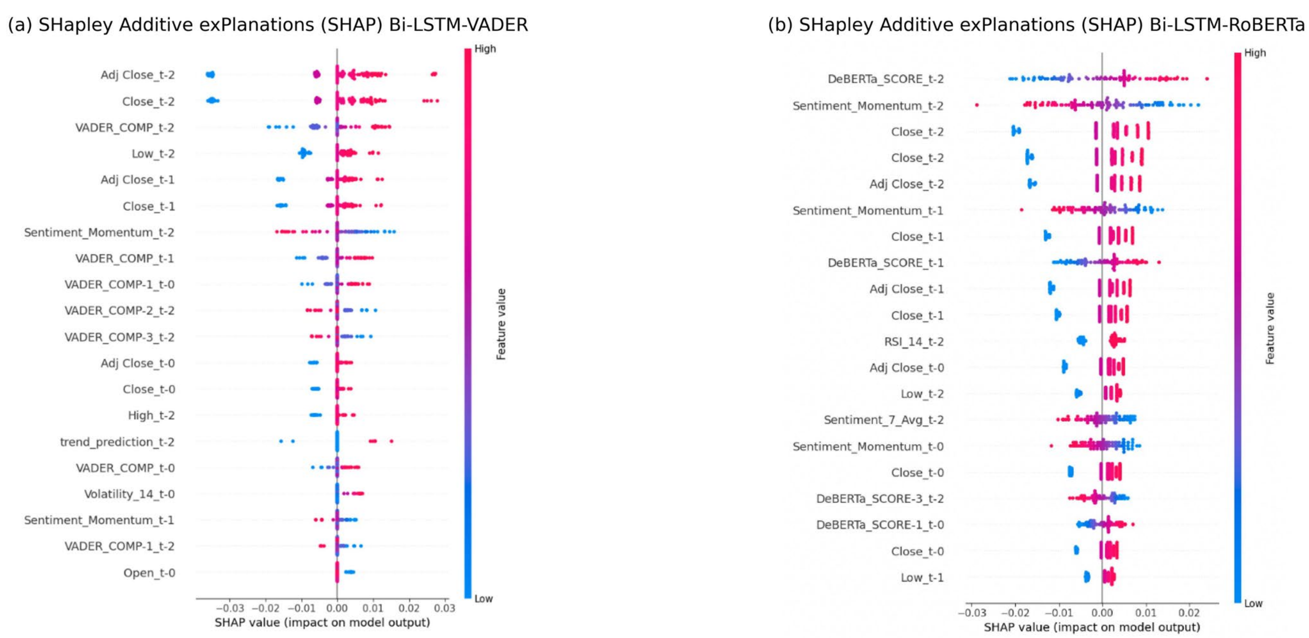

Fig. 13 SHAP feature contributions for Bi-LSTM Models Using VADER (**a**) and RoBERTa (**b**) Sentiment Analysis

model in sentiment-driven cryptocurrency forecasting is its ability to dynamically weigh historical price movements and sentiment trends, potentially yielding superior predictive accuracy compared to traditional recurrent models.

4.13 MAPE, RMSE, MAE, and directional accuracy results

The Temporal Attention Model was evaluated with both RoBERTa and VADER sentiment analysis models, using four key metrics: Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Directional Accuracy. The results are summarized in Table 7.

The model demonstrates strong absolute error reduction capabilities, with RoBERTa achieving a lower MAPE (2.26%) and RMSE (0.02) compared to VADER (MAPE = 3.30%, RMSE = 0.07). This suggests that contextual embeddings from RoBERTa provide a more refined sentiment representation, leading to improved forecasting accuracy. However, both models exhibit low Directional Accuracy (RoBERTa: 53%, VADER: 48%), indicating a significant limitation in predicting market trends correctly.

4.14 Temporal attention model (TAM) analysis of prediction accuracy: VADER vs. RoBERTa

The Temporal Attention Model (TAM) using VADER sentiment analysis, illustrated in Fig. 14a, shows enhanced capability in capturing cryptocurrency price trends compared to traditional recurrent models. The predicted price (red dashed line) aligns closely with the actual price (blue line) under stable market conditions, indicating the model's effectiveness in learning market dynamics. Nonetheless, during periods of heightened volatility, the predictions exhibit step-like patterns, highlighting persistent challenges in smoothly adjusting to abrupt market shifts. These inaccuracies may arise from limitations associated with VADER's lexicon-based sentiment scoring, which struggles to adapt dynamically to nuanced market sentiment changes.

In contrast, the TAM utilizing RoBERTa sentiment embeddings, as shown in Fig. 14b, demonstrates significantly improved alignment with actual prices, especially during periods of market stability and gradual trends. The closer tracking of predicted versus actual prices reflects the superior predictive power afforded by RoBERTa's contextualized sentiment embeddings. This is supported by enhanced performance metrics, including lower MAPE (2.26%), reduced RMSE (0.02), and lower MAE (0.01), confirming the accuracy and reliability of RoBERTa-enhanced TAM for cryptocurrency price forecasting. However, directional accuracy remains relatively modest (53%), suggesting ongoing challenges in consistently predicting price movement directions despite accurate absolute price estimations. Overall, these findings indicate that transformer-based, context-aware sentiment embeddings substantially improve TAM forecasting performance over lexicon-based approaches, although further refinements may be necessary to enhance directional prediction accuracy.

Table 7 Performance Metrics Comparison for Temporal Attention-Based Sentiment Models

Metric	TAM (RoBERTa)	TAM (VADER)
MAPE (%)	2.26	3.30
RMSE	0.02	0.07
MAE	0.01	0.08
Directional Accuracy (%)	53.0	48.0

4.15 Comparison of all model performance metrics

The comparison of performance metrics across LSTM, GRU, Bi-LSTM, and TAM models summarized in Table 8, reveals that Bi-LSTM (RoBERTa) remains the best-performing architecture in terms of MAPE (2.01%) and directional accuracy (79.5%), reinforcing the effectiveness of bidirectional sequence learning in financial forecasting. GRU

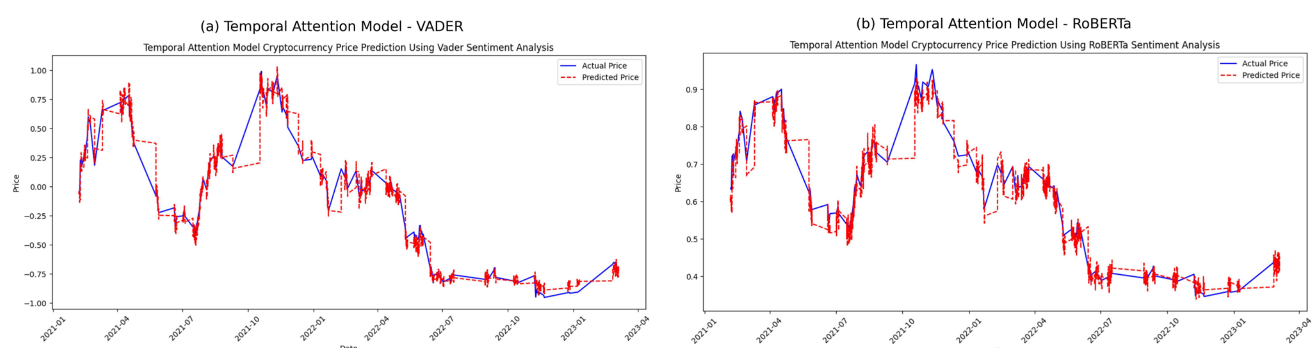

Fig. 14 Comparison of Actual vs. Predicted BTC Prices: (**a**) TAM-VADER Model, (**b**) TAM-RoBERTa Model

Table 8 Performance Metrics Comparison Across All Models

Model	MAPE (%)	RMSE	MAE	Directional accuracy (%)
LSTM (RoBERTa)	2.54	0.01	0.01	69.27
LSTM (VADER)	4.83	0.05	0.03	51.01
GRU (RoBERTa)	4.36	0.02	0.06	62.3
GRU (VADER)	5.5	0.08	0.05	50.02
Bi-LSTM (RoBERTa)	2.01	0.01	0.015	79.5
Bi-LSTM (VADER)	3.90	0.05	0.025	70.1
TAM (RoBERTa)	2.26	0.02	0.01	53.0
TAM (VADER)	3.30	0.07	0.08	48.0

exhibits the weakest performance, particularly with VADER (MAPE of 5.5% and directional accuracy of 50.02%), suggesting that it struggles to incorporate sentiment-driven signals effectively. LSTM demonstrates moderate performance, but it remains inferior to Bi-LSTM, indicating that bidirectional architectures offer a distinct advantage. The Temporal Attention Model (TAM) with RoBERTa achieves a strong MAPE (2.26%), comparable to Bi-LSTM but with significantly lower directional accuracy (53%), indicating that while it excels at absolute price prediction, it struggles with directional trend forecasting. TAM with VADER performs even worse, with higher MAPE (3.30%) and the lowest directional accuracy (48%) among all models, reaffirming that lexicon-based sentiment models are suboptimal for financial forecasting. The results further confirm that deep learning-based sentiment models (RoBERTa) consistently enhance predictive accuracy across all architectures, while attention-based architectures like TAM require additional optimizations to improve trend classification.

The SHAP analysis further highlights the impact of sentiment features on model performance. RoBERTa-based sentiment scores (DeBERTa_SCORE$_{t-2}$, Sentiment_Momentum$_{t-2}$) show strong predictive influence in Bi-LSTM, whereas in LSTM and GRU, historical price features remain dominant, with sentiment playing a minor role. This suggests that Bi-LSTM is the most effective at leveraging sentiment momentum for price forecasting, whereas GRU and LSTM rely more heavily on past price movements. Additionally, VADER-based models consistently underperform compared to RoBERTa-based models, reinforcing the notion that deep contextual embeddings capture financial sentiment dynamics more effectively than lexicon-based sentiment models. Overall, the findings confirm that both model architecture and the choice of sentiment analysis approach significantly impact cryptocurrency price prediction accuracy, with Bi-LSTM and RoBERTa emerging as the optimal combination for sentiment-driven forecasting.

Table 9 Comparison of Model Performance with Related Works

Time Period	Model	MAPE (%)
5 August 2022 - 5 September 2022	Bi-LSTM	12.32
	GRU	11.47
	FinBERT-NN	9.91
5 September 2022 - 5 October 2022	Bi-LSTM	12.01
	GRU	11.57
	FinBERT-NN	9.78
Average	Bi-LSTM	12.17
	GRU	11.52
	FinBERT-NN	9.85
28 February 2023 - 5 March 2023 (Our Approach)	LSTM (RoBERTa)	2.54
	LSTM (VADER)	4.83
	GRU (RoBERTa)	4.36
	GRU (VADER)	5.5
	Bi-LSTM (RoBERTa)	2.01
	Bi-LSTM (VADER)	3.90
	TAM (RoBERTa)	2.26
	TAM (VADER)	3.30

The SHAP analysis not only enhances the interpretability of the forecasting models but also offers practical value for quantitative analysts and algorithmic trading systems. The results highlight that sentiment related features, particularly those capturing momentum, and temporal interactions carry significant predictive weight. These insights can inform the development of sentiment-derived trading signals, guide dynamic portfolio adjustment during periods of sentiment shift, and support adaptive model calibration based on evolving feature importance patterns.

4.16 Comparison of all model performance metrics

See Table 8.

4.17 Comparing with related works

In Table 9, we compare the results of our models with the related work by Haritha and Sahana (2023) titled"Cryptocurrency Price Prediction Using Twitter Sentiment Analysis."Their study utilized Bi-LSTM, GRU, and FinBERT-NN models, and they reported Mean Absolute Percentage Error (MAPE) for both real-time and test data.

The study reported the following MAPE values for their models:

The results highlight the superior performance of the proposed sentiment-driven cryptocurrency forecasting models compared to previous works. In particular, Bi-LSTM with RoBERTa sentiment features achieves the lowest Mean Absolute Percentage Error (MAPE) of 2.01%, significantly outperforming the benchmark study, where the best-performing model, FinBERT-NN, achieved a MAPE of 9.85%. Similarly, other architectures such as LSTM (RoBERTa) and the Temporal Attention Model (TAM) with RoBERTa also demonstrate strong predictive capabilities, achieving MAPE values of 2.54% and 2.26%, respectively. The findings reinforce the importance of incorporating deep contextual embeddings in sentiment-driven financial forecasting, as models utilizing RoBERTa consistently outperform those relying on traditional lexicon-based sentiment analysis(VADER). Additionally, while the proposed models excel in absolute price prediction accuracy, directional accuracy varies, with Bi-LSTM achieving the highest directional accuracy of 79.5%. This suggests that bidirectional sequence learning is particularly effective in capturing sentiment-driven market movements. Overall, these results establish the proposed approach as a more robust framework for sentiment-enhanced cryptocurrency price forecasting compared to existing methodologies.

5 Conclusion

This study investigates the role of social media sentiment in cryptocurrency price forecasting by integrating deep learning models (LSTM, GRU, Bi-LSTM, and the Temporal Attention Model) with VADER and RoBERTa sentiment analysis. The findings demonstrate that sentiment-driven models significantly enhance price prediction accuracy, with Bi-LSTM (RoBERTa) achieving the lowest MAPE of 2.01%, outperforming traditional lexicon-based approaches. These results highlight the importance of deep contextual embeddings in financial sentiment modeling, reinforcing the value of alternative data sources in financial forecasting. The study also identifies key sentiment-driven features influencing price movements. SHAP analysis revealed that Sentiment Momentum, RoBERTa Compound Score, and VADER Negativity Score were the most impactful predictors, suggesting that investors react more to sentiment shifts than absolute sentiment values. Additionally, the high directional accuracy of Bi-LSTM (79.5%) indicates that bidirectional learning enhances market movement prediction. These findings emphasize the importance of sentiment momentum and polarity in understanding cryptocurrency price fluctuations.

The comparative analysis further highlights the superiority of deep contextual sentiment embeddings over lexicon-based sentiment methods. Models incorporating RoBERTa sentiment features consistently outperformed those using VADER, demonstrating the advantages of transformer-based sentiment extraction in capturing subtle market sentiment shifts. This reinforces the need for advanced sentiment analysis techniques in financial modeling and provides a framework for further enhancing cryptocurrency price forecasting.

5.1 Limitations and future work

While the results are promising, this study has some limitations. First, it focuses solely on Bitcoin, which may limit the generalizability of findings to other cryptocurrencies. Additionally, real-time implementation remains computationally expensive, particularly for RoBERTa-based sentiment analysis, which required over six hours on high-performance GPUs (2x RTX 4090). This makes scalability a challenge for real-time trading applications. The computational performance of RoBERTa-based sentiment analysis presents a significant challenge for real-time implementation, as running sentiment extraction alone on high-performance hardware is time-intensive, making real-time inference impractical for high-frequency trading. Future optimizations should explore lightweight transformer models (e.g., DistilBERT, TinyBERT), caching mechanisms for precomputed sentiment scores, and hybrid approaches where VADER handles real-time sentiment shifts while RoBERTa provides deeper batch-mode insights. Additionally, deploying these models in a low-latency environment (e.g., via ONNX, TensorRT, or quantized models) could improve feasibility for trading applications. Moreover, Twitter sentiment analysis is susceptible to bot-generated content, fake news, and market manipulation, which could bias the training data. Implementing robust data filtering techniques and combining multiple sentiment sources (e.g., Reddit, news articles) could help mitigate these biases.

Future research should explore faster sentiment models, expand the dataset to multiple cryptocurrencies, and incorporate additional financial indicators to enhance predictive accuracy. Further investigation is also needed to assess the long-term viability of sentiment-based trading strategies and their implications for market behavior.

Funding Open access funding provided by Sefako Makgatho Health Sciences University. Open access funding provided by Sefako Makgatho Health Sciences University. Data will be made available by the authors on request.

Data availability Data will be made available by the authors on request.

Declarations

Conflict of interest The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

References

Aliber RZ, Kindleberger CP, McCauley RN (2023) Bitcoin: Worse than a ponzi. In *Manias, panics, and crashes: a history of financial crises*, pages 349–371. Springer

Antonopoulos AM (2017) *Mastering bitcoin: programming the open blockchain*. " O'Reilly Media, Inc."

Arias M, Arratia A, Xuriguera R (2014) Forecasting with twitter data. ACM Trans Intell Syst Technol (TIST) 5(1):1–24

Baker M, Wurgler J (2006) Investor sentiment and the cross-section of stock returns. J Finance 61(4):1645–1680

Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. J Computat Sci 2(1):1–8

Bouoiyour J, Selmi R (2019) How do futures contracts affect bitcoin prices? Econom Bull 39(2):1127–1134

Bouri E, Molnár P, Azzi G, Roubaud D, Hagfors LI (2017) On the hedge and safe haven properties of bitcoin: is it really more than a diversifier? Finance Res Lett 20:192–198

Bouri E, Gil-Alana LA, Gupta R, Roubaud D (2019) Modelling long memory volatility in the bitcoin market: evidence of persistence and structural breaks. Int J Finance Econom 24(1):412–426

Briere M, Oosterlinck K, Szafarz A (2015) Virtual currency, tangible return: portfolio diversification with bitcoin. J Asset Manag 16:365–373

Chainalysis Team. How does bitcoin's fixed supply impact its price?, 2023. URL https://www.chainalysis.com/blog/money-supply/. Accessed: 2024-06-11

Cheah E-T, Fry J (2015) Speculative bubbles in bitcoin markets? An empirical investigation into the fundamental value of bitcoin. Econom Lett 130:32–36

Cho K, Van Merriënboer B, Bahdanau D, Bengio Y (2014a) On the properties of neural machine translation: encoder-decoder approaches. *arXiv preprint* arXiv:1409.1259

Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y (2014b) Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint* arXiv:1406.1078

Ciaian P, Rajcaniova M, Kancs A (2016) The economics of bitcoin price formation. Appl Econom 48(19):1799–1815

Ciaian P, Rajcaniova M et al (2018) Virtual relationships: short-and long-run evidence from bitcoin and altcoin markets. J Int Financ Mark Inst Money 52:173–195

Colianni S, Rosales S, Signorotti M (2015) Algorithmic trading of cryptocurrency based on twitter sentiment analysis. CS229 Project 1(5):1–4

Corbet S, Lucey B, Urquhart A, Yarovaya L (2019) Cryptocurrencies as a financial asset: a systematic analysis. Int Rev Financ Anal 62:182–199

Cunha J, Murphy C (2019) Are cryptocurrencies a good investment? J Invest 28(3):45–56

Duan J, Wang P, Ma W, Tian X, Fang S, Cheng Y, Chang Y, Liu H (2021) Short-term wind power forecasting using the hybrid model of improved variational mode decomposition and correntropy long short-term memory neural network. Energy 214:118980

Effrosynidis D, Arampatzis A, Symeonidis S (2018) Sentiment analysis of tweets using a lexicon-based approach. In *Proceedings of the 8th International Conference on Management of Digital EcoSystems*, pages 168–175. ACM

Fama EF (1970) Efficient capital markets: a review of theory and empirical work. J Finance 25(2):383–417

Fischer T, Krauss C (2018) Deep learning with long short-term memory networks for financial market predictions. Europ J Operat Res 270(2):654–669

Gang-Jin W, Chi X, Danyan W, Longfeng Z (2019) When bitcoin meets economic policy uncertainty (epu): measuring risk spillover effect from epu to bitcoin. Finance Res Lett 31:14

Gao Y, Wang R, Zhou E (2021) Stock prediction based on optimized lstm and gru models. Sci Progr 2021(1):4055281

Garcia D, Schweitzer F (2015) Social signals and algorithmic trading of bitcoin. Royal Soc Open Sci 2(9):150288

Griffin JM, Shams A (2018) Manipulation in the vix? Rev Financ Stud 31(4):1377–1417

Haritha G, Sahana N (2023) Cryptocurrency price prediction using twitter sentiment analysis. In *CS & IT conference proceedings*, volume 13. CS & IT Conference Proceedings

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Computat 9(8):1735–1780

Howell ST, Niessner M, Yermack D (2020) Initial coin offerings: Financing growth with cryptocurrency token sales. Rev Financ Stud 33(9):3925–3974

Hutto CJ, Gilbert E(2014a) Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*, pages 216–225

Hutto C, Gilbert E (2014) Vader: A parsimonious rule-based model for sentiment analysis of social media text. Proc Int AAAI Conf Web Soc Media 8:216–225

Ider D, Lessmann S (2022) Forecasting cryptocurrency returns from sentiment signals: An analysis of bert classifiers and weak supervision. *arXiv preprint* arXiv:2204.05781

Janze C (2017) Are cryptocurrencies criminals best friends? examining the co-evolution of bitcoin and darknet markets

Jothi Prakash V, Arul Antran Vijay S (2024) A multi-aspect framework for explainable sentiment analysis. Patt Recogn Lett 178:122–129

Jim Kyung-Soo Liew and Garrett Zhengyuan Wang (2016) Twitter sentiment and ipo performance: a cross-sectional examination. J Portfol Manag 42(4):129

Katsiampa P (2017) Volatility estimation for bitcoin: a comparison of garch models. Econom Lett 158:3–6

Khadjeh NA, Saeed A, Ying WT, David CLN (2015) Text mining of news-headlines for forex market prediction: a multi-layer dimension reduction algorithm with semantics and sentiment. Expert Syst Appl 42(1):306–324

Kraaijeveld O, De Smedt J (2020) The predictive power of public twitter sentiment for forecasting cryptocurrency prices. J Int Financ Mark Inst Money 65:101188

Krouska A, Troussas C, Virvou M (2016) The effect of preprocessing techniques on twitter sentiment analysis. In *2016 20th Pan-Hellenic Conference on Informatics (PCI)*, pages 1–6. IEEE

Lahmiri S, Bekiros S (2019) Cryptocurrency forecasting with deep learning chaotic neural networks. Chaos Solit Fractals 118:35–40

Lee J, Park J, Kim Keunhyoung L, Nam J (2017) Sample-level deep convolutional neural networks for music auto-tagging using raw waveforms. *arXiv preprint* arXiv:1703.01789

Lim B, Zohren S (2021) Time-series forecasting with deep learning: a survey. Philosoph Trans Royal Soc A 379(2194):20200209
Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, Stoyanov V (2019) Roberta: a robustly optimized bert pretraining approach. *arXiv preprint* arXiv:1907.11692
Loughran T, McDonald B (2011) When is a liability not a liability? Textual analysis, dictionaries, and 10-ks. J Finance 66(1):35–65
Mai F, Shan Z, Bai Q, Wang X, Chiang RHL (2018) How does social media impact bitcoin value? a test of the silent majority hypothesis. J Manag Inform Syst 35(1):19–52
McNally Sean, Roche Jason, Caton Simon (2018) Predicting the price of bitcoin using machine learning. In *2018 26th euromicro international conference on parallel, distributed and network-based processing (PDP)*, pages 339–343. IEEE
Medhat W, Hassan A, Korashy H (2014) Sentiment analysis algorithms and applications: a survey. Ain Shams Eng J 5(4):1093–1113
Mittal A, Goel A (2012) Stock prediction using twitter sentiment analysis. *Standford University*, CS229 (2011) http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf, 15:2352
Moin Amani, ün Sirer Emin G, Sekniqi Kevin (2019) A classification framework for stablecoin designs. *arXiv preprint* arXiv:1910.10098
Murugesapandian M (2023) Sentiment analysis on cryptocurrency tweets using machine learning. Int J Sci Res Eng Manag 07:1–11. https://doi.org/10.55041/IJSREM25924
Nakamoto Satoshi (2008) Bitcoin: A peer-to-peer electronic cash system. *Decentralized business review*
Narayanan Arvind, Bonneau Joseph, Felten Edward, Miller Andrew, Goldfeder Steven (2016) *Bitcoin and cryptocurrency technologies: a comprehensive introduction*. Princeton University Press
Neil Gandal JT, Hamrick TM, Oberman T (2018) Price manipulation in the bitcoin ecosystem. J Monet Econom 95:86–96
Pang B, Lee L et al (2008) Opinion mining and sentiment analysis. Foundat Trends 2(1–2):1–135
Pano T, Kashef R (2020) A complete vader-based sentiment analysis of bitcoin (btc) tweets during the era of covid-19. Big Data Cognit Comput 4(4):33
Patel J, Shah S, Thakkar P, Kotecha K (2015) Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. Expert Syst Appl 42(1):259–268
Paul A, Nayyar A et al (2024) A context-sensitive multi-tier deep learning framework for multimodal sentiment analysis. Multimed Tools Appl 83(18):54249–54278
Phillips RC, Gorse D (2017) Predicting cryptocurrency price bubbles using social media data and epidemic modelling. In *2017 IEEE symposium series on computational intelligence (SSCI)*, pages 1–7. IEEE
Raheman A, Kolonin A, Fridkins I, Ansari I, Vishwas M (2022) Social media sentiment analysis for cryptocurrency market prediction. *arXiv preprint* arXiv:2204.10185
Rognone L, Hyde S, Sarah Zhang S (2020) News sentiment in the cryptocurrency market: an empirical comparison with forex. Int Rev Financ Anal 69:101462
Schuster M, Paliwal KK (1997) Bidirectional recurrent neural networks. IEEE Trans Signal Proc 45(11):2673–2681
Shen D, Urquhart A, Wang P (2019) Does twitter predict bitcoin? Econom Lett 174:118–122
Shiller RJ (2003) From efficient markets theory to behavioral finance. J Econom Perspect 17(1):83–104
Sompolinsky Y, Zohar A (2015) Secure high-rate transaction processing in bitcoin. In *Financial cryptography and data security: 19th international conference, FC 2015, San Juan, Puerto Rico, January 26-30, 2015, Revised Selected Papers 19*, pp 507–527. Springer
Sovbetov Y (2018) Factors influencing cryptocurrency prices: evidence from bitcoin, ethereum, dash, litcoin, and monero. J Econom Financ Anal 2(2):1–27
Suresh Kaushik. Bitcoin tweets. https://www.kaggle.com/datasets/kaushiksuresh147/bitcoin-tweets, 2023. Accessed: 2024-06-11
SVA Labs. twitter-xlm-roberta-bitcoin-sentiment. https://huggingface.co/svalabs/twitter-xlm-roberta-bitcoin-sentiment. [Accessed 2023-03-07]
Symeonidis S, Effrosynidis D, Arampatzis A (2018) A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis. Expert Syst Appl 110:298–310
Taboada M, Brooke J, Tofiloski M, Voll K, Stede M (2011) Lexicon-based methods for sentiment analysis. Computat Linguist 37(2):267–307
Tetlock PC (2007) Giving content to investor sentiment: the role of media in the stock market. J Finan 62(3):1139–1168
Urquhart A (2017) Price clustering in bitcoin. Econom Lett 159:145–148
Valencia F, Gómez-Espinosa A, Valdés-Aguirre B (2019) Price movement prediction of cryptocurrencies using sentiment analysis and machine learning. Entropy 21(6):589
Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Łukasz PI (2017) Attention is all you need. Adv Neural Inform Process Syst 30:102458
Vigna Paul, Casey Michael J (2016) *The age of cryptocurrency: how bitcoin and the blockchain are challenging the global economic order*. Macmillan
Vitalik B et al (2014) A next-generation smart contract and decentralized application platform. White Paper 3(37):2–11
Werner Sam, Perez Daniel, Gudgeon Lewis, Klages-Mundt Ariah, Harz Dominik, Knottenbelt William ((2022)) Sok: Decentralized finance (defi). In *proceedings of the 4th ACM conference on advances in financial technologies*, pages 30–46
Wood G et al (2014) Ethereum: a secure decentralised generalised transaction ledger. Ethereum Project Yellow Paper 151(2014):1–32
Wanshan W, Tiwari AK, Gozgor G, Leping H (2021) Does economic policy uncertainty affect cryptocurrency markets? evidence from twitter-based uncertainty measures. Res Int Bus Finan 58:101478
Xiao J, Deng T, Bi S (2024) Comparative analysis of lstm, gru, and transformer models for stock price prediction. In: *proceedings of the international conference on digital economy, blockchain and artificial intelligence*, pp 103–108
Yermack David (2015) Is bitcoin a real currency? an economic appraisal. In *Handbook of digital currency*, pp 31–43. Elsevier
Yong Yu, Si X, Changhua H, Zhang J (2019) A review of recurrent neural networks: Lstm cells and network architectures. Neural Computat 31(7):1235–1270
Zerveas G, Jayaraman S, Patel D, Bhamidipaty A, Eickhoff C (2021) A transformer-based framework for multivariate time series representation learning. In: *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pages 2114–2124
Zhang L, Wang S, Liu B (2018) Deep learning for sentiment analysis: a survey. Wiley Interdiscipl Rev Data Mining Knowl Disc 8(4):e1253
Zhao J, Gui X (2017) Comparison research on text pre-processing methods on twitter sentiment analysis. In *2017 IEEE international conference on big data (BIGDATA)*, pp 2870–2879. IEEE
Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, Zhang W (2021) Informer: Beyond efficient transformer for long sequence time-series forecasting. Proc AAAI Conf Artif Intell 35:11106–11115
Zimbra D, Abbasi A, Zeng D, Chen H (2018) The state-of-the-art in twitter sentiment analysis: a review and benchmark evaluation. ACM Trans Manag Inform Syst (TMIS) 9(2):5:1-5:30